Mechanical Systems and Signal Processing 229 (2025) 112554

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

On the influence of attributes for assessing similarity and sharing knowledge in heterogeneous populations of structures

Giulia Delo [a,*], Aidan J. Hughes [b], Cecilia Surace [c], Keith Worden [b]

[a] Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino, 10129, Italy
[b] Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK
[c] Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino, 10129, Italy

ARTICLE INFO

Communicated by O. Fink

Keywords:
Population-based structural health monitoring (PBSHM)
Transfer learning
Domain adaptation
Joint distribution adaptation (JDA)
Graph matching networks

ABSTRACT

The effectiveness of data-driven Structural Health Monitoring methods, which require dynamic response data from undamaged and damaged states within the same domain, is often hindered by the scarcity of data. The population-based approach to Structural Health Monitoring (PBSHM) aims to overcome this challenge by exploiting the transfer of damage-state knowledge across a population of structures via transfer-learning algorithms — specifically domain adaptation. However, meaningful inference through data sharing is possible only when the structures and their datasets are sufficiently similar. This study advances PBSHM by focusing on its two main phases: assessing structural similarity and developing effective transfer-learning strategies for heterogeneous populations. Furthermore, it explores the interdependence of these phases by addressing the role of structural attributes – such as material properties, geometry, and dimensions – in the performance of domain adaptation and similarity metrics. The proposed methodology employs Normal Condition Alignment (NCA) and Joint Distribution Adaptation (JDA) as domain-adaptation techniques across a population of laboratory-scale aircraft, considering tasks of increasing complexity. Additionally, the effect of material properties, geometry or dimensional differences in similarity assessment is investigated, comparing Graph Matching Networks and correlation-based distance metrics.

1. Introduction

Structural Health Monitoring (SHM) is a crucial strategy for ensuring safety and reducing the maintenance costs of structures across various fields. With the advent of advanced sensor technologies and data analytics, there has been a shift towards data-driven methods, such as machine learning and pattern recognition [1]. These methods leverage vibration-based data from multiple sensors to enhance real-time assessments, detect anomalies, and predict potential issues, offering more accurate and cost-effective decision-making. However, the experimental data for training machine-learning methods are often unavailable or incomplete. Thus, the practical implementation of these methods is limited, and most machine-learning methods for SHM have been validated only in laboratory conditions [2–4]. Indeed, these innovative predictive models would require extensive measurements of a structure's dynamic response in multiple environmental and operational conditions. Moreover, if a supervised method is implemented, it would also be necessary to have labelled data available from the structure in damaged conditions [1]. Population-based Structural Health Monitoring (PBSHM) has been introduced as a novel approach for improving the performance of data-driven SHM [5–7]; it aims to overcome the issues of data scarcity by exploiting transfer-learning algorithms for sharing damage-state knowledge across a

* Corresponding author.
E-mail address: giulia.delo@polito.it (G. Delo).

https://doi.org/10.1016/j.ymssp.2025.112554
Received 30 August 2024; Received in revised form 20 January 2025; Accepted 4 March 2025

population of structures. Transfer learning leverages knowledge gained from a source domain, where data is available, to enhance learning in a related but distinct target domain with limited data. In the context of PBSHM, this involves identifying similarities between the source and target structures and their damage-sensitive features, and using domain adaptation techniques to align their distributions, enabling effective knowledge transfer.

Thus, it is possible to exploit the damage-sensitive information of a source structure with a comprehensive dataset to improve the diagnostic performance of a different target structure with only a few available data.

PBSHM distinguishes *homogeneous* and *heterogeneous* populations. Homogeneous populations include only nominally-identical systems or structures; i.e., structures in which only minor differences exist, such as those related to the manufacturing process. These populations are the focus of [5], which opens the sequence dedicated to introducing the foundations of the PBSHM theory. Heterogeneous populations, instead, include structures which can have more differences from each other and can exhibit distinct behaviours, as highlighted in [6]. As homogeneous structures exhibit analogous behaviours, sharing knowledge among them may be straightforward. In contrast, considering a heterogeneous population, such a strategy would not always lead to satisfactory performance. Sharing knowledge between two heterogeneous structures could enhance the performance if the source and target domains are sufficiently similar. Otherwise, using the source domain in the training process could lead to the *negative transfer* phenomenon. This phenomenon consists of failing to properly harmonise the two domains via transfer-learning strategies and mislabelling classes, as defined in [7], and could lead to worse overall performance than using only the target structure data.

As a consequence, to limit the risk of negative transfer, PBSHM requires the introduction of a preliminary phase before transfer learning, but equally relevant, namely similarity assessment. Assessing similarity across the members of a population helps to define between which structures or sub-components it would be possible to apply knowledge-transfer strategies and which features could be considered, according to the specific monitoring task and the type of diagnosis required, i.e., damage detection, localisation, or quantification [8]. Afterwards, the effectiveness of transfer learning, computed as the expected value of information transfer (EVIT) [9], can provide valuable insights for making informed decisions regarding operational processes, maintenance protocols, and optimal PBSHM strategies for new target structures.

The issue of negative transfer induced by high dissimilarity between the source and the target domains has been considered in a few studies, including [10,11], which proved the need to introduce similarity assessment. Similarity assessment has been addressed using the notion of task relatedness, as proposed in [12], and task clustering [13]. Many distance metrics adopted in this context have been reviewed in [14]. In addition, [15] recently investigated the implementation of distance metrics to evaluate the similarity between features, including the Hellinger, the total variational, and the Area metrics, and their combination. However, within the PBSHM, an automatic similarity assessment should likewise be possible on the structures, before analysing the measured dynamic characteristics. Therefore, structures have been represented abstractly in the form of Irreducible Element (IE) models and Attributed Graphs (AGs) [6] and distance functions for graph-based representations have been investigated.

The IE model of a structure is defined as a simplified representation, which describes in the form of elements only the main structural components, i.e., components having a well-established dynamic behaviour, such as wings, fuselage, tail, plates or beams. In addition, the elements are connected using specific relationships to model joints and boundary conditions, and it is possible to associate attributes with each element of the model, which encompass characteristics such as element type, geometry, materials and dimensions. These models can have different levels of granularity, according to the detail adopted for component subdivision, which depends on the SHM task addressed, and the IE complexity rapidly increases for real-world structures. Thus, a Canonical Form (CF) IE model has been introduced in [16] to facilitate automated comparisons between IE models built with different granularity. The IE models are subsequently transformed into AGs using a graph structure in which elements are represented as nodes and relationships are represented as edges. Moreover, nodes and edges can have attributes to store the initial IE properties. These graph structures are used to perform comparisons via graph-based algorithms and quantitatively evaluate a similarity index.

Distance metrics for graph-structured data are categorised into methods based on graph matching, edit distance, refinement graphs, or kernels, according to [17]. An early approach proposed in [6] to measure similarity involves calculating the maximum common subgraph (MCS) and the Jaccard Index. This metric falls in the class of graph-matching methods, and it consists of a relaxation of graph isomorphisms [18]. The MCS identifies the largest subset of common nodes and edges between two structures AGs. The relative dimension of this subgraph with respect to the graph's size is used to compute a similarity index. Nevertheless, the performance of the MCS is limited by variations in AG granularity, and the use of CF models would be required. Moreover, this approach is not convenient to consider multiple attributes per node or edge, or multiple types of attributes, such as discrete and continuous ones. Given the challenges with the MCS-based method and the Jaccard index, subsequent research efforts focused on alternative methods for similarity assessment, such as kernel-based methods [19]. The random-walk graph kernel has been implemented in [20] for assessing structural similarity. This approach derives from a previous study by Borgwardt et al., in which graph kernels were proposed for finding proteins with similar characteristics [21]. It employs the calculation of a convolutional kernel to identify all possible equivalent random walks in the two AGs. The main advantage of this method is the opportunity to choose a proper kernel function for handling different types of attributes, e.g., discrete and continuous ones. Other approaches for measuring structural similarity are based on graph refinement [22] and on graph edit distances, which simulate structural variation using edit operations and measure the distance between two graphs based on the minimum sequence of edit operations needed to transform a graph into the second one, as shown in [23]. An interesting application of graph edit distance is shown in [24], where this method is compared to a fingerprint-based one to measure molecular similarity. However, these approaches may have high computational complexity [17]. Therefore, further research focused on more efficient graph edit distance algorithms suitable if the data is represented as a tree [25].

Li et al. [26] proposed a similarity assessment method based on graph matching networks (GMNs), an extension of graph neural networks (GNN). GMNs leverage a cross-graph attention mechanism to match nodes across AGs, increasing the accuracy of the similarity scores. This methodology has been implemented in the PBSHM framework for civil structures, mainly focussing on graph topology. Subsequently, numerical and categorical attributes have been embedded as vectors for each node and edge of the graphs, and element and relationship types have been integrated into a bridge case study in [16]. The GMN-based similarity assessment, though, has been exploited by embedding only a few attributes specific to the IE scheme. Thereby, structure-specific properties were not considered in the similarity index, although these properties may play a crucial role in the performance of PBSHM. Most notably in heterogeneous populations, the main differences may be caused by material properties, geometry or dimensions, so considering them in similarity assessment becomes essential [27]. An initial analysis of how to embed such attributes in the GMN similarity rules was presented in [28]. However, it is still necessary to understand how to implement these distance metrics, considering geometrical and material differences.

The interest towards transfer-learning strategies for SHM emerged because of the limitations in implementing many machine-learning approaches for real-world structures. Transfer learning is a branch of machine learning suitable for solving identification, classification and regression problems on target systems having limited training data [10,29]. This class of algorithms includes different approaches to extract knowledge from a source system and share it with a new target system to improve and generalise a learning task. The main differences in the algorithms are related to the homogeneity or heterogeneity of the tasks and the availability of labelled data in the training dataset of the target system [30]. A few research studies have already adopted transfer-learning techniques for SHM, using deep convolutional neural networks (DCNN) on acquired images, especially for the classification and detection of cracks [31,32], and also for the investigation of fatigue damage [33]. Furthermore, domain adaptation [34], a branch of transfer learning, proved particularly well suited for PBSHM purposes because of its ability to harmonise the source and target datasets and facilitate the different levels of vibration-based structural damage identification. Gardner et al. proposed the application of domain adaptation for PBSHM [35], implementing the Transfer Component analysis (TCA), the Joint Distribution Adaptation (JDA), and the Adaptation regularisation-based transfer learning for damped natural frequencies and damping ratios of numerical and experimental laboratory-scale three-storey structures. Some applications of unsupervised transfer learning have been proposed in the context of bridge monitoring, as shown in [36–38], and for anomaly detection in electric motor bearings and jet engines by exploiting adversarial deep learning [39]. Moreover, domain adaptation has been introduced for guided wave SHM [40], and Zhou et al. proposed a fuzzy-set-based joint distribution adaptation for improving damage quantification on a helicopter panel [41]. Subsequently, the Balanced Distribution Adaptation (BDA) algorithm has been implemented for transferring knowledge between two aircraft wings, using transmissibility functions as features [42], and TCA has been validated on an experimental dataset comprising Frequency Response Functions (FRFs) acquired from three different tail planes [43]. Further applications of TCA are detailed in [44], in which this strategy is applied to experimental frame structures, and in [45], where the TCA is combined with an inverse Finite Element methodology (iFEM) to handle measurements from various sources. In addition, a statistic alignment (SA) strategy, already proposed as a pre-processing step in [42], has been exploited as a low-risk form of domain adaptation [46] in a numerical case study, and a real case study involving two monitored bridges.

However, the choice of transfer-learning method, that is -"how to transfer" [30]- is still challenging and depends significantly on what features are available for knowledge-sharing. Furthermore, "when to transfer" remains a fundamental research issue for designing a transfer-learning algorithm, and for PBSHM, it can be addressed through the similarity-evaluation phase. Therefore, it is necessary to analyse the relationship between the two phases and the effect of the attribute on the performance of the domain-adaptation methods. The current research addresses this need by analysing the variation in domain-adaptation performance on a heterogeneous population of laboratory-scale aircraft. In addition, this paper aims to present a discussion on appropriate similarity metrics. Accordingly, a similarity assessment based on GMNs is implemented, drawing upon a similarity metric developed for a population of synthetic bridges, as shown in [16], extending this approach to consider attribute influences. Subsequently, the presented approach is compared to a correlation-motivated distance metric to further stress the value of physical and mechanical properties in discerning heterogeneous structures. The overall flow diagram of the proposed research process is depicted in 1. Two damage-identification tasks are implemented via domain-adaptation algorithms for each pair of experimental structures to harmonise the modal features, transforming them into a low-dimensional latent space where damage can be detected, as preliminarily investigated in [47]. Confusion matrices are employed to compare the findings and show how these properties can influence the transfer-learning performance for different levels of task complexity, thus highlighting the importance of their evaluation in the context of PBSHM.

The layout of the paper is as follows. Section 2 presents the design and construction of the heterogeneous population of experimental aircraft structures, performed to obtain a comprehensive dataset for studying and validating PBSHM approaches. Section 3 introduces the domain-adaptation algorithms used for sharing knowledge within the population, and Section 4 presents the transfer-learning results. Section 5 describes the GMN distance metric and its results. Subsequently, Section 6 introduces a correlation-based similarity assessment and shows the corresponding outcomes. Lastly, Section 7 presents the discussions and some conclusions.

2. A population of experimental aircraft structures

The current section describes a heterogeneous population of laboratory-scale aircraft models, focussing on the structures' design and the experimental setup. The population includes eight similar models. Each one has been designed according to a benchmark study to obtain highly simplified and economical structures, yet able to simulate the dynamic behaviour of an aircraft. The reference

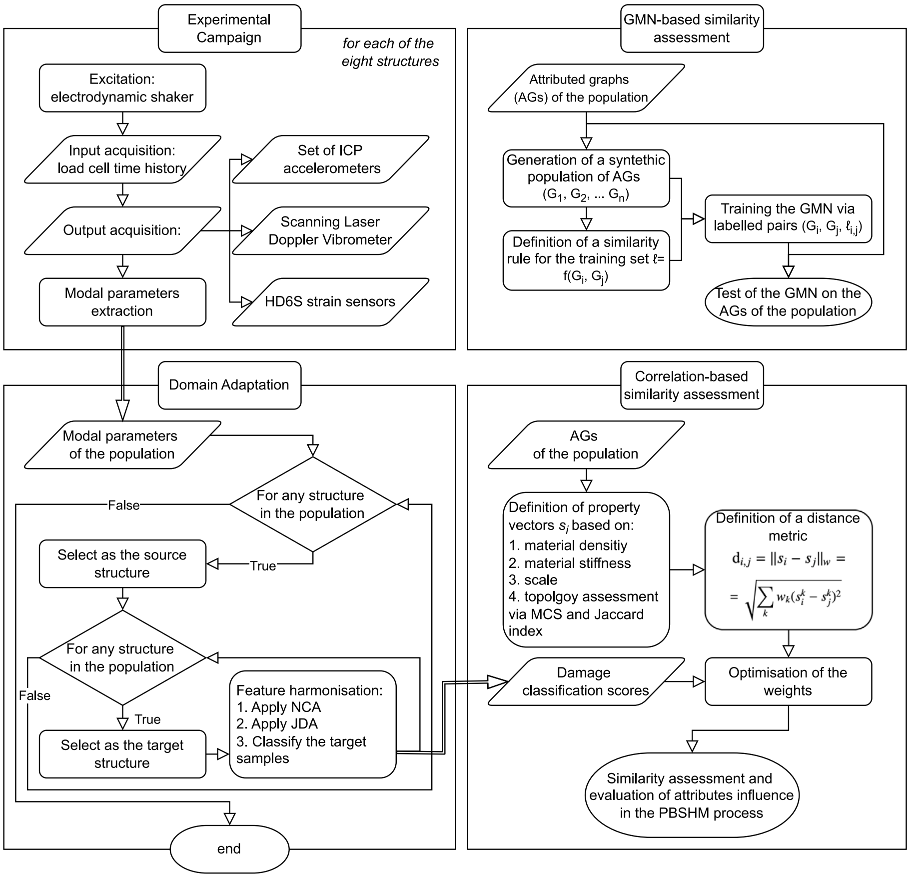
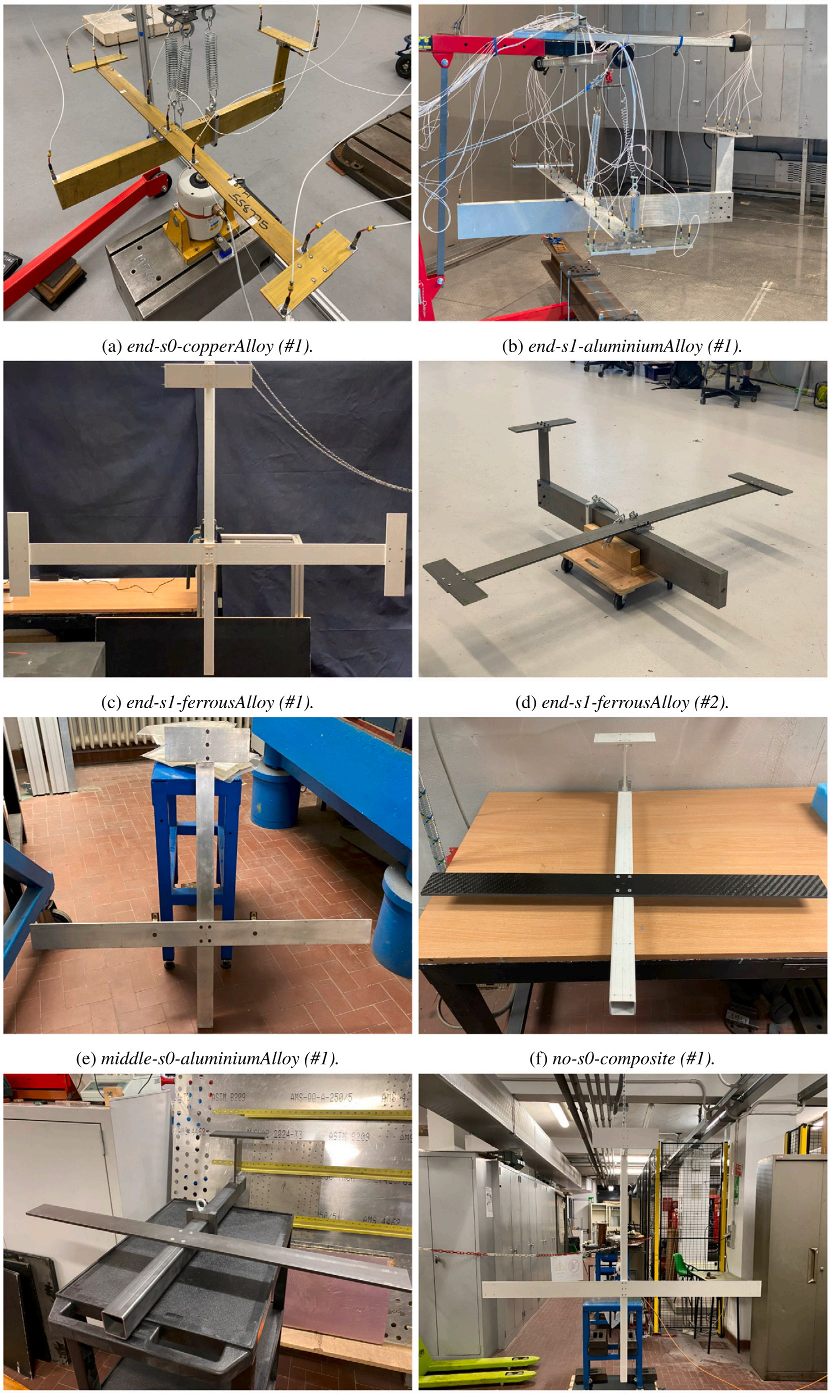


Fig. 1. Flow diagram of the proposed approach, including the experimental campaign, the domain adaptation analysis, and the two similarity assessment strategies.

experimental campaign has been conducted by the Structures and Materials Action Group (SM-AG19) of the Group for Aeronautical Research Technology in EURope (GARTEUR) [48,49]. The GARTEUR study involved different research groups to investigate the approaches for ground vibration tests of aircraft and compare the performance of multiple methods for extracting the modal properties. Therefore, several copies of the structure had been produced, resulting in a homogeneous population. These models comprise beam and plate components joined together via bolted connections. The fuselage is a thin-section rectangular beam, and it is connected to a single rectangular plate representing the wings. Similar plate components are connected to the fuselage to create the vertical and the horizontal tail. The new aircraft models are characterised by some variations in geometric properties, such as the sections of the various components, dimensions, and materials used, to introduce heterogeneity in the population [50], resulting in a comprehensive experimental dataset of heterogeneous structures. The test structures are shown in Fig. 2 sorted according to their topology, i.e., with end winglets, no winglets or middle engines. For simplicity, each structure is identified by the name highlighted in the Figure. Further details on the topology classes can be found in Section 5. The aircraft are distinguished according to two scaling classes (s0 and s1). Structures (b, c, d and h) in Fig. 2 have a wingspan of 2.0 m. These further differ because of the materials used, i.e., steel or aluminium alloys, and different fuselage geometries. The second scaling class includes structures (a, e, f and g) in Fig. 2. These aircraft have a wingspan of 1.0 m. In addition, aircraft (a) is made of copper alloy, aircraft (e) is made of aluminium alloy, aircraft (f) has steel fuselage and tail, but its wings are made of composite materials, and aircraft (g) is made of ferrous alloy (specifically steel).

The structures have been tested via classical and strain Experimental Modal Analyses (EMA) to compare different outputs and features. Five models (a, e, f, g and h) have been tested via periodic chirp excitation and Scanning Laser Doppler Vibrometer (SLDV). The other three have been tested via random-noise excitation and accelerometers, as in the original study. In addition, strains have been acquired from two models (a and c). In either case, a single input has been imposed using a shaker connected to the rear part of the fuselage, and a suspension system has been used to simulate free vibration conditions. For a more comprehensive description of the performed strain EMA, the reader is directed to [51]. The tests have been carried out in the “LAQ-AERMEC Aeromechanical Structural Systems” laboratory of the Department of Mechanical and Aerospace Engineering, Politecnico di Torino, and the “Laboratory for Verification and Validation” (LVV), The University of Sheffield. The current study considers one undamaged



(a) *end-s0-copperAlloy (#1).* (b) *end-s1-aluminiumAlloy (#1).*

(c) *end-s1-ferrousAlloy (#1).* (d) *end-s1-ferrousAlloy (#2).*

(e) *middle-s0-aluminiumAlloy (#1).* (f) *no-s0-composite (#1).*

(g) *no-s0-ferrousAlloy (#1).* (h) *no-s1-ferrousAlloy (#1).*

Fig. 2. Pictures of the heterogeneous population of aircraft structures.

condition and six different damaged conditions, which have been acquired at ambient temperature. Damage has been simulated by applying concentrated masses (around 2% of the aircraft mass) in multiple positions over the structure: one at the wing tip, two along the wing, one on the horizontal tail and two on the fuselage, at the tip and the end. The decision to use mass addition as a proxy for damage was guided by the need to create a reversible and tunable variation in structural response that could be analysed in the PBSHM framework, as suggested in [52,53]. The mass effects on the modal properties have been replicated in a finite-element model

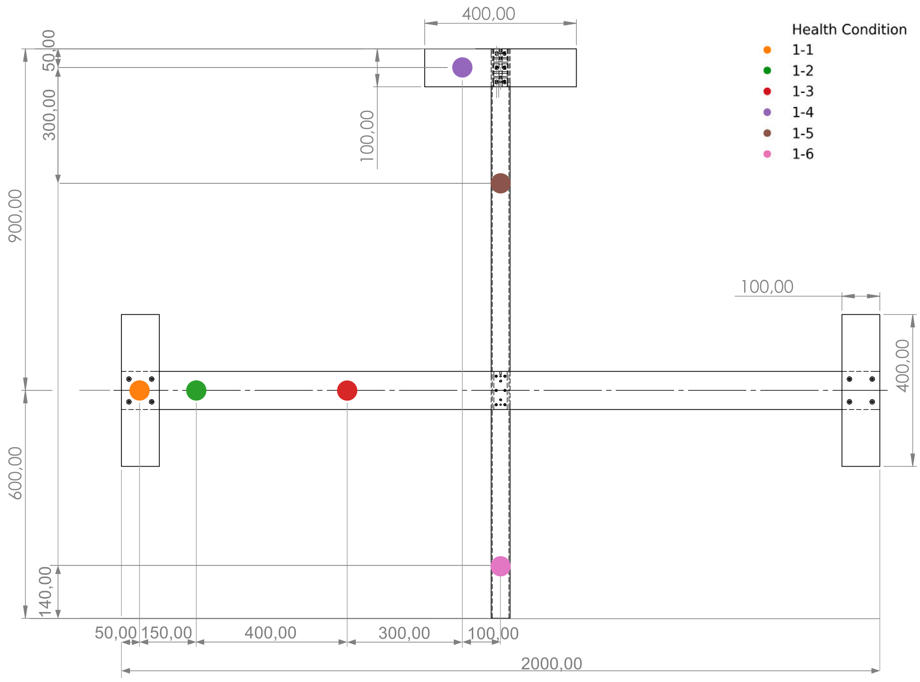


Fig. 3. Top view of the large aircraft scheme in the topological configuration with the winglets showing the location of the concentrated masses used to simulate the six damaged conditions. In the health conditions legend, the first number is 1 if the condition is labelled as damaged, 0 otherwise; the second number is the damage location index.

Table 1
Description of source and target analysed conditions.

Label	Presence of pseudo-faults	Mass position
0-0	No	–
1-1	Yes	Winglet
1-2	Yes	Wing tip
1-3	Yes	Centre of the wing
1-4	Yes	Fuselage tip
1-5	Yes	Rear part of the fuselage
1-6	Yes	Tail

of the structures by introducing a 50% reduction of the Young's modulus in specified regions to define an equivalent additional mass. Although this approach does not capture the full complexity of real damage scenarios, it provides a well-defined and measurable way to introduce variability in the system conditions, suitable for validating the methodology presented in the study. The positions of the concentrated masses are shown in Fig. 3 for the large aircraft with winglets (#1), and their description is included in Table 1. However, these conditions are consistently re-scaled across the population. The responses have been acquired in the form of FRFs at multiple locations; this provided the input to be analysed, from which various damage-sensitive features could be extracted, depending on the specific problem to be addressed. The FRFs have been analysed over 2048 spectral lines, and the frequency range has been set within 320 Hz for the smaller-scale models and 80 Hz for the larger-scale models. The modal properties have been analysed by adopting the pyEMA Python library [54]. The current study adopts the extracted natural frequencies, reported in Table A.3 as input features for the damage-identification tasks.

3. Domain adaptation

Employing transfer-learning approaches to leverage damage-sensitive knowledge from similar structures can enhance diagnostic performance, as demonstrated experimentally in [43] for a population of three tailplanes and in [55] for three of the laboratory-scale aircraft shown in Section 2. In the current study, the knowledge transfer is implemented for the complete aircraft population by examining all the different combinations of source domain and target domain. Two different tasks are considered: damage detection and damage localisation. Thus, it is possible to gain insight into how the differences in the analysed structures influence the transfer-learning process on different tasks. Among the different transfer-learning approaches available, this study focusses on JDA, which falls inside the class of domain-adaptation algorithms [34].

To introduce these algorithms, the following definitions must be considered, as stated in [30]. A domain $\mathcal{D}$ consists of a feature space χ, and a marginal probability distribution $p(X)$, where $X = \{x_1, x_2, \dots, x_n\} \in \chi$ are the data samples. A task $\mathcal{T}$ consists of a label space $\mathcal{Y}$, where $Y = \{y_1, y_2, \dots, y_n\} \in \mathcal{Y}$, and a predictive function $p(y_i|x_i)$, which can be learnt from the training process and subsequently used to make label predictions on new data samples. For instance, these features could be portions of a transfer function such as an FRF or the vector of identified natural frequencies, while the labels could be indicators of damage ($Y = \{0, 1\}$), its location ($Y = \{0, 1, \dots, C\}$), or its intensity (in this case, the problem could be defined as a classification problem with discrete intensity labels, or as a regression problem with continuous labels). Transfer-learning algorithms aim at improving the predictive function on a target domain $\mathcal{D}_T$ using the source domain $\mathcal{D}_S$ and task $\mathcal{T}_S$, assuming that $\mathcal{D}_T \neq \mathcal{D}_S$ or $\mathcal{T}_T \neq \mathcal{T}_S$. The current study deals with transductive feature-based transfer learning, aiming to harmonise source and target features using a common representation to enhance diagnostic inferences on the target structure [30]. Specifically, transductive transfer learning assumes that $\mathcal{T}_T = \mathcal{T}_S$ but $\mathcal{D}_T \neq \mathcal{D}_S$, and the domain difference can concern the features ($\chi_S \neq \chi_T$), or their marginal probability distributions ($p(X_S) \neq p(X_T)$) [10]. In these cases, many labelled data are required in the source domain, and no labelled data are required in the target domain. The latter can be traced back to domain adaptation because, in this class of algorithms, the difference between the source and the target domains is given by the marginal probability distributions [56]. As in the transductive setting tasks are equivalent, it is assumed that $p(Y_S|X_S) = p(Y_T|X_T)$. Although this assumption can be used in domain adaptation, it may not apply to some knowledge-sharing problems [34]. Therefore, some domain-adaptation algorithms, such as TCA [34] and JDA [57], weaken this assumption and consider different conditional probability distributions.

JDA has been introduced by Long et al. [57] as a feature-based method that maps the source and target features onto a latent space with reduced dimensionality via a non-linear mapping to a Reproducing Kernel Hilbert Space (RKHS) $\psi : \chi \rightarrow H$, where the difference between the marginal probability distributions and the difference between the conditional probability distributions are minimised: $p(\psi(X_S)) \approx p(\psi(X_T))$ and $p(Y_S|\psi(X_S)) \approx p(Y_T|\psi(X_T))$. Thus, the transformed features can be used to train a classifier for solving the task on the target domain. The distance between the probability distributions can be measured using the non-parametric Maximum Mean Discrepancy (MMD) criterion [58] with an empirical kernel embedding, as shown in Eq. (1) [35].

$$\mathrm{Dist}(p(\psi(X_S)), p(\psi(X_T))) + \mathrm{Dist}(p(Y_S|\psi(X_S)), p(Y_T|\psi(X_T))) \approx tr(W^T K M_C K W) \tag{1}$$

Given a kernel function $k(x_i, x_j) = \psi(x_i)^T \psi(x_j)$, K is the kernel matrix for X, $X = (X_S + X_T) \in \mathbb{R}^{\mathrm{N}x\mathrm{D}}$ is the union of the source and target features used during training, $\mathrm{N} = (\mathrm{N}_S + \mathrm{N}_T)$ is the number of training samples from the two domains, and D is their initial dimension. This study implemented a Radial Basis Function (RBF) kernel, which was selected to ensure smooth mappings and better separation of the data clusters,

$$k(x_i, x_j) = \exp\left(\frac{\|x_i - x_j\|}{2l^2}\right) \tag{2}$$

where l is the length-scale and can be estimated using the median heuristic [59]. In addition, $W \in \mathbb{R}^{\mathrm{N}x\mathrm{k}}$ is the matrix performing the feature transformations and reducing their dimension to k, and M_C is the class conditional MMD matrix, where $c = \{0, 1, \dots, C\}$ are the possible classes. M_C [57] is expressed as,

$$M_C(i,j) = \begin{cases} \frac{1}{\mathrm{N}_S^{(c)}\mathrm{N}_S^{(c)}}, & \text{if } x_i, x_j \in \mathcal{D}_S^{(c)} \\ \frac{1}{\mathrm{N}_T^{(c)}\mathrm{N}_T^{(c)}}, & \text{if } x_i, x_j \in \mathcal{D}_T^{(c)} \\ \frac{-1}{\mathrm{N}_S^{(c)}\mathrm{N}_T^{(c)}}, & \text{if } x_i \in \mathcal{D}_S^{(c)},\ x_j \in \mathcal{D}_T^{(c)} \\ \frac{-1}{\mathrm{N}_T^{(c)}\mathrm{N}_S^{(c)}}, & \text{if } x_i \in \mathcal{D}_T^{(c)},\ x_j \in \mathcal{D}_S^{(c)} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

In the expression for M_C, $\mathcal{D}_S^{(c)}$ is the set of source samples belonging to class c, while $\mathrm{N}_S^{(c)}$ is the number of source samples in class c. Analogously, for the target domain, $\mathcal{D}_T^{(c)}$ is the set of samples belonging to class c, while $\mathrm{N}_T^{(c)}$ is the number of samples in class c. The current study assumes that there are no available labelled samples in the target domain for each of the classes, which would be required to compute M_C. Thus, the training samples of the target domains are classified using pseudo-labels, which are computed using a classifier trained on the source data. The KNN classifier [60] is adopted in this study. The optimisation problem can be written as an eigenvalue problem, where W consists of the eigenvectors corresponding to the k smallest eigenvalues,

$$\left(K \sum_{c=0}^{C} M_c K + \lambda I\right) W = K H K W \psi \tag{4}$$

In Eq. (4), $H = \mathcal{I}_{\mathrm{NxN}} - (1/\mathrm{N}) \cdot [1]_{\mathrm{NxN}}$. Because of the pseudo-labelling process, the optimisation must be repeated iteratively to refine the target labels and the transformation matrix. Thus, in addition to k and the regularisation parameter λ, the maximum number of iterations becomes a hyperparameter of the algorithm. Having derived the definitive transformation matrix, the transformed features $Z = KW \in \mathbb{R}^{\mathrm{N}x\mathrm{k}}$, can be used to train the KNN classifier on the target task.

This approach is developed for the experimental population data, adopting the first seven natural frequencies identified from classical EMA as input features. An unsupervised standardisation is performed so that each feature component is treated equally [61], and it is applied individually to each domain. Features normalisation is a critical step in any machine-learning strategies and for facilitating domain adaptation, as highlighted in [43]. The preprocessing pipeline involves a normalisation based on the z-score strategy, which is appropriate for the experimental data clusters, as they exhibit Gaussian characteristics. In the context of domain adaptation, this unsupervised preprocessing strategy can potentially align the first-order statistics of the marginal probability distributions, thereby mitigating the risk of negative transfer. The source dataset includes labelled data. Instead, it is assumed that, within the target dataset, there are labelled data corresponding solely to samples in an undamaged condition. This assumption allows for the inclusion of an additional pre-processing step, which can enhance the efficiency and effectiveness of the overall process. Before the application of JDA, the features are subjected to Normal Condition Alignment (NCA), a Statistic Alignment (SA) approach described in [46]. The NCA computes the means and standard deviations from the undamaged condition samples in the source and target datasets (i.e., $\mu_{s,n}, \sigma_{s,n}$ and $\mu_{t,n}, \sigma_{t,n}$), and applies a transformation to the target dataset, which aligns its undamaged condition samples to the source undamaged condition (Eq. (5)). The NCA step would be highly beneficial in cases of class-imbalance between the source and the target domains [46].

$$z_t^{(j)} = \frac{x_t^{(j)} - \mu_{t,n}}{\sigma_{t,n}} \cdot \sigma_{s,n} + \mu_{s,n} \tag{5}$$

The first analysed task regards damage detection. Thus the labels are $Y = \{0, 1\}$. The second task regards damage localisation. In the latter, three damage locations are considered. i.e., wings, fuselage and tail. Therefore, the labels are $y = 0$ for samples in the undamaged class, $y = 1$ if the samples are related to damage conditions on the wings, $y = 2$ if the samples are related to damage conditions on the fuselage, and $y = 3$ if the samples are related to damage conditions on the tail. In addition, the damage-detection and localisation performance after domain adaptation is compared to the one gathered without employing the PBSHM strategy. In particular, the kNN classifier trained on the source data is used to label the target data after normalisation without taking into account differences between source and target domains.

4. Domain-adaptation results

The current section shows the domain-adaptation results computed on the heterogeneous population of experimental aircraft. The two SHM tasks are evaluated on each possible combination of structures. Thus, the overall transfer-learning performance is shown via confusion matrices, in which the accuracy and the F1 score are reported for each pair of source and target domains. Additionally, it is possible to gain further insights into the performance of each specific knowledge-transfer problem by analysing the transformed features and the KNN confusion matrices, which compare the actual and the predicted target labels. Since the experimental tests lead to limited samples for each health condition, the extracted features are re-sampled from an uncertain representation to extend the dataset, ensuring sufficient variability for the knowledge-sharing application. This representation is obtained by considering the mean experimental features and polluting them with additive Gaussian noise ($\sigma = 0.1$). The training and test datasets are built as follows. The training and test datasets are defined to study a knowledge-sharing scenario in which the source is a data-rich structure for which all the samples are labelled. Instead, the target is a data-poor structure for which fewer training samples are available. Only samples related to the target normal condition are labelled, while the others are unlabelled. This condition arises from the assumption that a few normal-condition data are collected from the target structure at the beginning of its operational life, facilitating the statistical alignment of features before further domain adaptation steps. Based on these assumptions, the training and testing datasets can be delineated as follows. However, further studies could also explore different scenarios of incomplete target data. In addition, the definition of data-rich structures is not univocal, as the amount and type of data required for successful knowledge transfer can vary depending on the complexity of the target task and the adopted domain adaptation approach. Therefore, extensions of this study could evaluate thresholds to determine the dimensions and completeness required to obtain an appropriate source dataset.

- The training dataset from the source structure comprises 50 labelled samples from the undamaged condition and 50 labelled samples from each damaged condition.
- The training dataset from the target structure comprises 25 labelled samples from the undamaged condition and 25 unlabelled samples from each damaged condition.
- The test dataset from the target structure comprises 25 unlabelled samples from the undamaged condition and 25 unlabelled samples from each damaged condition.

In addition, the previously-described approach is repeated ten times, and the averaged performance is computed to reduce the influence of the sampling process on the findings. The JDA parameters are consistently set across the different tasks. The dimensionality of the JDA latent space, k, is set to 2 based on a combination of empirical observations and the characteristics of the datasets used in this study. Setting k to a lower dimension facilitates the learned latent space visualisation while maintaining sufficient representational capacity for damage identification tasks. The regularisation parameter, λ, is set to 0.1 to balance the trade-off between aligning the marginal and conditional distributions and avoiding overfitting, as suggested in the literature [35] and the maximum number of iterations used to compute the pseudo-labels on the target domain is set to 5 based on convergence analysis.

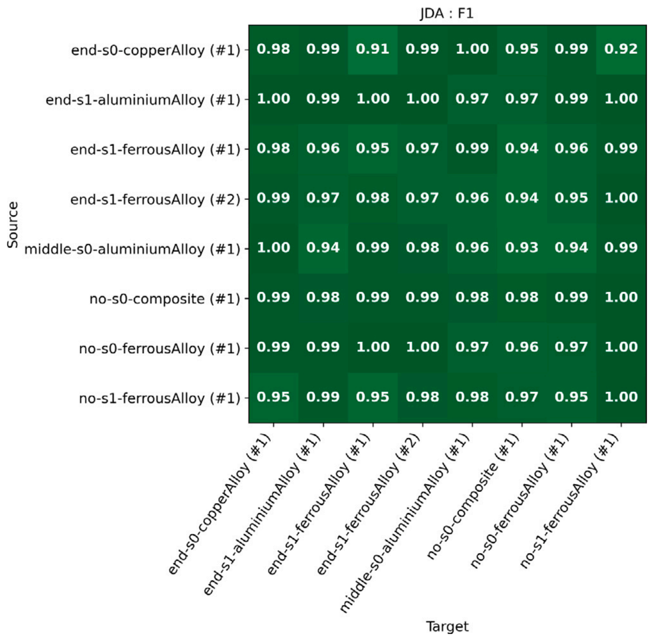


Fig. 4. Damage-detection performance across the population: F1 score confusion matrix.

4.1. Damage-detection task

The damage-detection task is performed by considering a binary classification problem. The labels are reduced to $Y = \{0, 1\}$, where 0 indicates an undamaged condition, and 1 indicates the presence of any of the six damages considered. The source and target natural frequencies are standardised and aligned via NCA. Subsequently, the JDA is adopted to harmonise the features, assuming that the information about damage presence is known only in the source domain. The adopted performance indicators are accuracy (6), precision (7), recall (8) and the F1 score (9). The overall results of the binary KNN classification are shown in Fig. B.18 in terms of accuracy, and Fig. 4 in terms of F1 score.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{6}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{F1score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

where TP is the number of true positive samples, i.e., correctly identified as undamaged; TN is the number of true negative samples, i.e., correctly identified as damaged; FP is the number of false positive samples, i.e., incorrectly identified as undamaged; FN is the number of false negative samples, i.e., incorrectly identified as damaged.

The results are discussed comparatively with respect to the various combinations of structures. These findings show excellent detection performance in most combinations, with an F1 score in the range $[0.91, 1]$. The accuracy and F1 values along the diagonals concern the cases in which knowledge is shared between two equivalent datasets, where the differences between the domains are induced by noise only, and the most successful results are expected. Conversely, the off-diagonal values indicate instances of heterogeneous transfer where additional differences exist. Despite the differences between the source and target structures, positive transfer is accomplished. The results were expected because all the structures originate from the same benchmark and demonstrate comparable dynamic behaviour. In addition, the task of distinguishing damaged or undamaged conditions appears to be straightforward, especially after NCA.

Each combination of structures is further analysed by considering how the different steps of the proposed approach, described in Section 3, transform the features and affect the classification. For the purpose of conciseness, the following analyses one combination in which the end-s1-ferrousAlloy (#1) GARTEUR is the source domain and the end-s1-aluminiumAlloy(#1) GARTEUR is the target domain. Fig. 5 displays the first three standardised natural frequencies before NCA, and Fig. 6 shows the same features after NCA. (The original input includes the first seven natural frequencies, whereas the figure shows only the first three to aid visualisation). The implication of the NCA step consists of the possibility of aligning the normal-condition samples using a low-risk form of domain

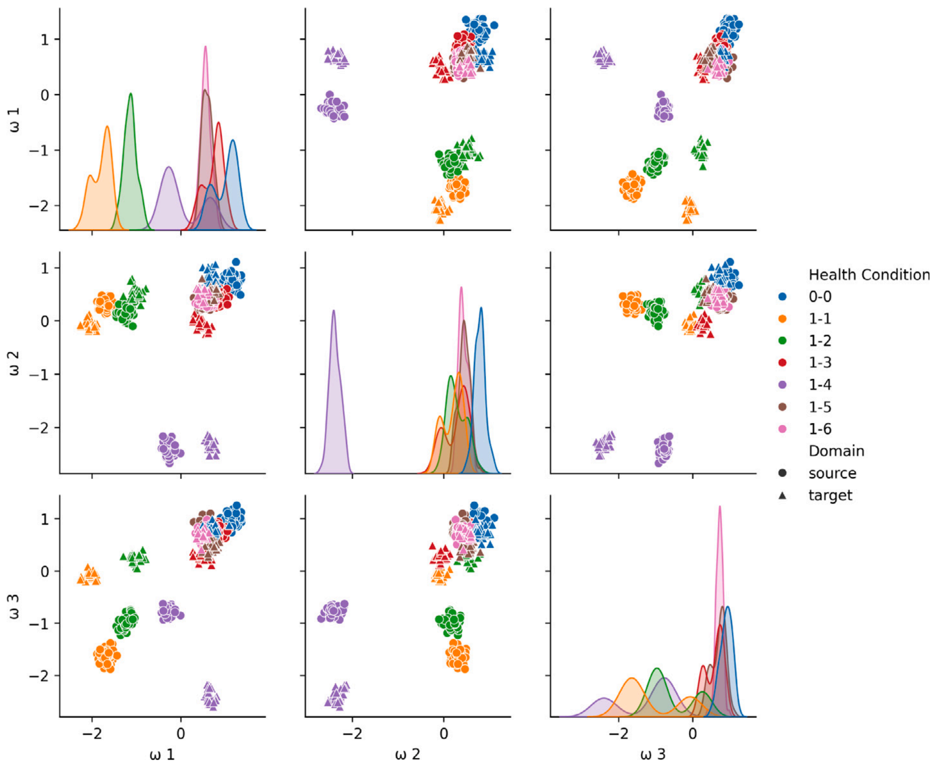


Fig. 5. First three natural frequencies from the end-s1-ferrousAlloy (#1) GARTEUR (source), and the end-s1-aluminiumAlloy(#1) GARTEUR (target) in the different experimental health conditions, before NCA.

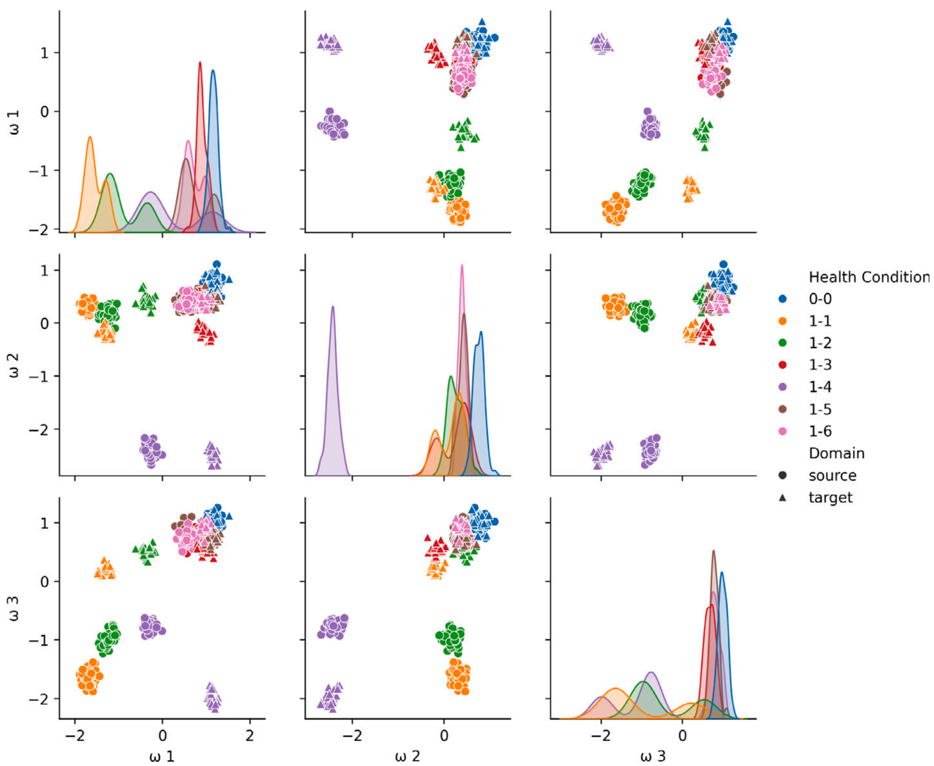


Fig. 6. First three natural frequencies from the end-s1-ferrousAlloy (#1) GARTEUR (source), and the end-s1-aluminiumAlloy(#1) GARTEUR (target) in the different experimental health conditions, after NCA.

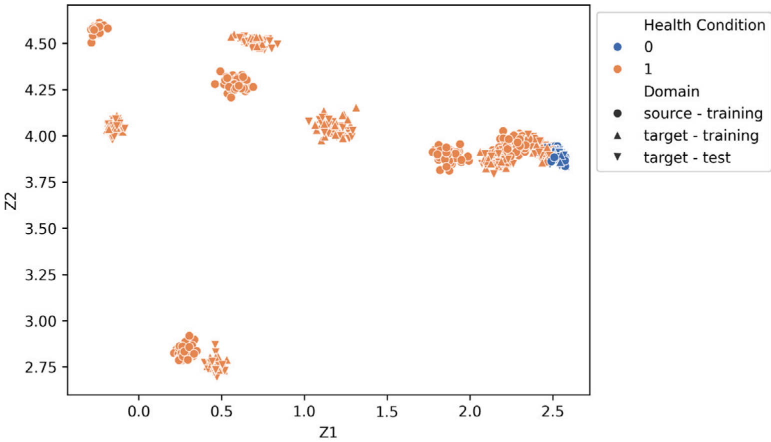


Fig. 7. Transformed features in the latent space $Z = \in \mathbb{R}^{N x k}$, where N is the number of samples and k = 2. These features are computed for the damage-detection task considering JDA on the end-s1-ferrousAlloy (#1) GARTEUR (source domain) and the end-s1-aluminiumAlloy(#1) GARTEUR (target domain).

adaptation by only assuming that a few normal-condition data can be collected from the target structure at the beginning of its operational life. Indeed, the NCA perfectly aligns the samples from the undamaged conditions (health condition: 0-0). However, the damaged-condition clusters are not aligned. In addition, these clusters are not well separated from the undamaged condition, suggesting the need for further domain adaptation. The JDA on the binary classification task leads to the bi-dimensional latent space shown in Fig. 7. Most of the samples in the damaged condition are well separated from the undamaged condition cluster, leading to a damage-detection accuracy of 93% and an F1 score of 96%. Figs. 5–7 are introduced to visually describe the effect of the two harmonisation steps, namely the NCA and the JDA. Although NCA is necessary for facilitating the second step of domain adaptation, the differences induced by the first transformation are lower than the ones produced by the JDA, as the NCA does not change the data structure, and the samples in Fig. 4 are already self-normalised. These results are compared to the ones concerning the damage detection performed without domain adaptation of each pair of source and target structures. In this case, the performance is considerably weaker, excluding the results on the diagonal of the confusion matrices, where source and target coincide. Indeed, the average F1 score is 30.8%, and the average accuracy is 52.8%. The confusion matrix concerning the task without domain adaptation is given in Appendix C (Fig. C.20). These results justify the need for domain-adaptation techniques to exploit data from different sources.

4.2. Damage-localisation task

The damage-localisation task is performed by considering a multinomial classification problem. As stated in Section 3, this task considered three damage locations. Thus, labels are reduced to $Y = \{0, 1, 2, 3\}$, where 0 indicates an undamaged condition; 1, 2 and 3 specify the presence of damage in the different main components, respectively, in the wings, the tail and the fuselage. After the pre-processing steps, which are analogous to the first task, the JDA is adopted, and the KNN is used with a One vs. Rest approach [61] to perform the four-class classification. The overall results are shown in Fig. B.19 in terms of accuracy, and Fig. 8 in terms of F1 score. By comparing Figs. B.18 and B.19, it can be observed that the localisation task is more significantly affected by the sources of heterogeneity in the aircraft population. Indeed, even if excellent detection results are obtained, the differences in topology, materials and dimensions lead to decreased performance in distinguishing the effects of damage in different positions. More evidently, the performance is maximum along the diagonal, while it tends to reduce when the differences between the aircraft increase. As expected, optimal results are also derived when sharing knowledge between the two aircraft having the same attributes, i.e., end-s1-ferrousAlloy (#1) and end-s1-ferrousAlloy (#2). Instead, the accuracy slightly decreases when the previous models are paired with the end-s1-aluminiumAlloy (#1), showing the influence of material attributes. The end-s0-copperAlloy aircraft (#1) has more dissimilar behaviour than other aircraft in the population. Therefore, when considered as a source or target structure, this structure leads to positive results if the second structure shares the same topology, otherwise lower results are obtained. The subset of aircraft with no winglets topological configuration also leads to good results, especially between no-s0-ferrousAlloy (#1) and no-s0-composite (#1), where the only difference is the wing material. By contrast, the accuracy derived using these two structures decreases when they are paired with the no-s1-ferrousAlloy aircraft (#1) because of the differences in dimensions. It is also interesting to analyse the behaviour of the latest aluminium aircraft. This structure comprises some elements to simulate the weight of the engines, leading to a different topology. However, this structure can be paired with discrete results with the other steel and composite aircraft having equivalent dimensions. Moreover, it leads to optimal results when paired with the two steel models having larger dimensions and end-winglets.

Analogously to the damage-detection task, these results are compared to the ones acquired without domain adaptation of each pair of source and target structures. A poor performance can be observed, excluding the results on the diagonal of the confusion

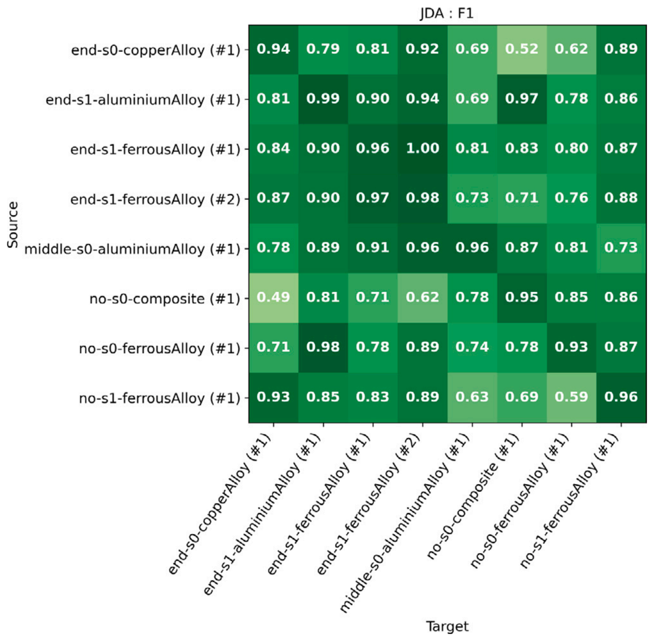


Fig. 8. Damage-localisation performance across the population: F1 score confusion matrix.

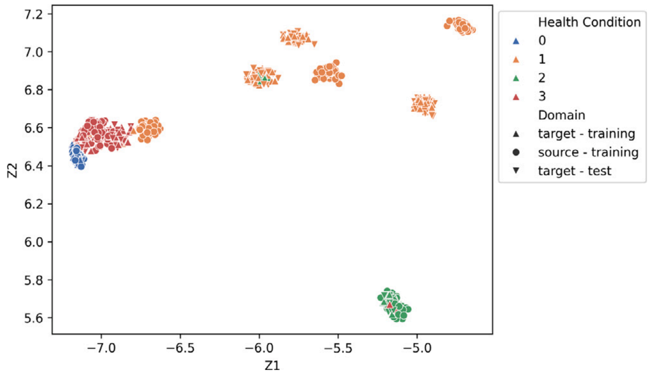


Fig. 9. Transformed features in the latent space $Z = \in \mathbb{R}^{N x k}$, where N is the number of samples and k = 2. These features are computed for the damage-localisation task considering JDA on the end-s1-ferrousAlloy (#1) GARTEUR (source domain) and the end-s1-aluminiumAlloy (#1) GARTEUR (target domain).

matrices, where source and target coincide. This indicates how domain-adaptation techniques are even more necessary for more complex tasks such as localisation. Indeed, the average F1 score is 11.1% and the average accuracy is 29.6%. The confusion matrix concerning the task without domain adaptation is given in Appendix C (Fig. C.21).

A detailed analysis of the transformed features and how this affects the KNN performance becomes particularly interesting for the damage-localisation task. The same combination of structures seen for the damage-detection task is considered. The pre-processing steps are analogous to the previous task. In this case, the JDA leads to the bi-dimensional latent space shown in Fig. 9. The samples concerning damage on the wings ($Y = 1$), and damage on the tail ($Y = 2$) are well aligned, leading to almost perfect performance. However, some target samples concerning damage on the fuselage ($Y = 3$) fall into the undamaged cluster, leading to their misclassification because the simulated damage on the fuselage, which is stiffer than the wings, has a minor impact on the dynamic behaviour of the structure. This outcome can be further clarified by Fig. 10, showing the KNN confusion matrix normalised with respect to the true labels.

The analysed combination of structures shows how even small differences, such as material attributes, lead to variations in dynamic behaviour and damage's effect. These variations may have limited influence on the first task, though their influence becomes more relevant in the second task, where some classes are not properly harmonised. In addition, when considering structures with higher attribute variations, the performance tends to decrease. This behaviour is shown in Figs. D.22 and D.23 (Appendix D), which correspond to the transfer-learning problem between the no-s1-ferrousAlloy (#1) GARTEUR and the middle-s0-aluminiumAlloy (#1) GARTEUR. This structure's combination is characterised by variations in topology, materials and dimensions. Thus, even if

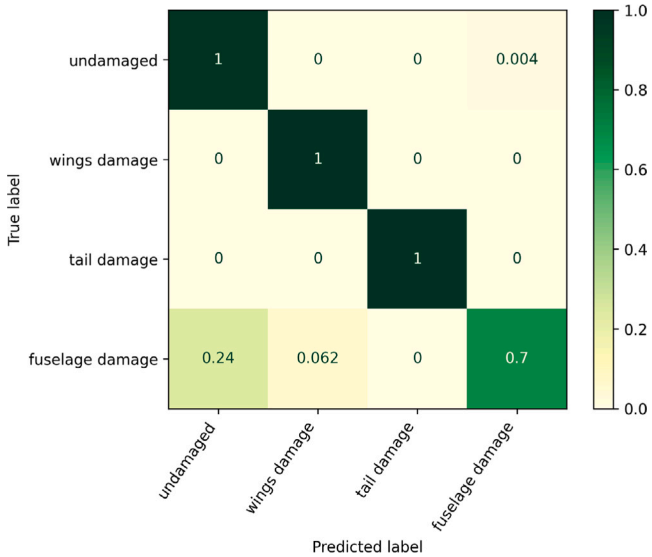


Fig. 10. KNN confusion matrix for the damage-localisation task considering JDA on the end-s1-ferrousAlloy (#1) GARTEUR (source domain) and the end-s1-aluminiumAlloy (#1) GARTEUR (target domain).

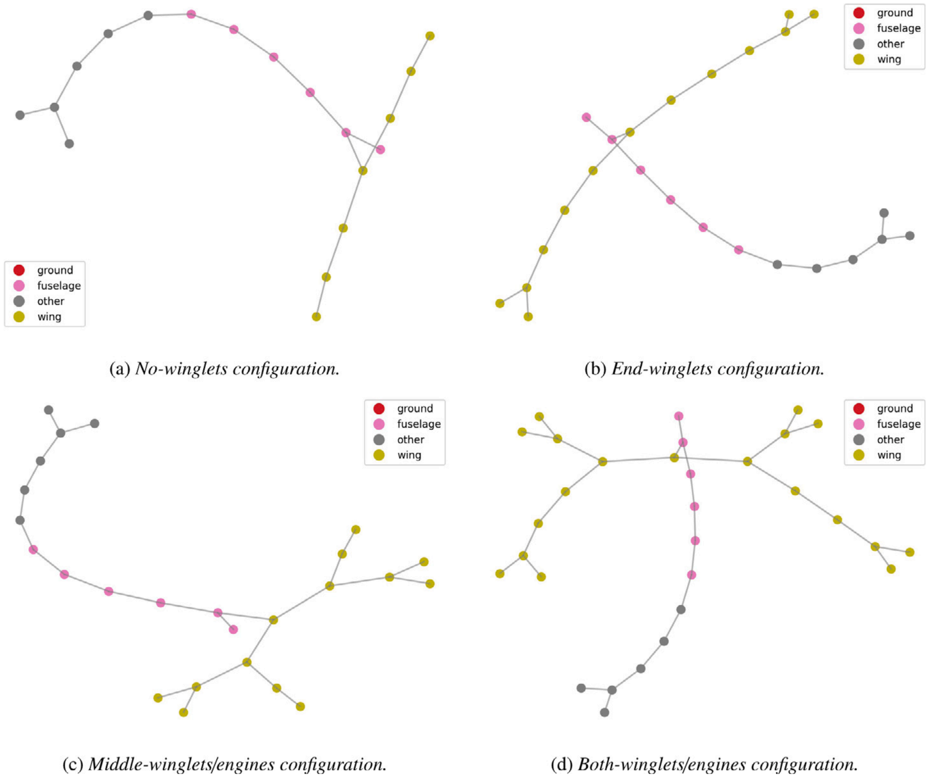
this combination provides good detection performance, the damage classes are incorrectly harmonised, producing 66% localisation accuracy and 63% F1 score.

5. GMN-based similarity assessment

Structural similarity can be assessed using different strategies, according to which structures are being compared and which parameters are most relevant for the specific case studies. If the SHM task involves detecting global damage, the similarity assessment should be primarily based on the topological aspects. Otherwise, if more localised effects must be analysed, or the task involves higher levels in the damage-identification hierarchy [8], considering some attributes becomes fundamental. The transfer learning process is influenced by the similarity between the type and nature of damage in the source and target domains, and it benefits significantly when the damage observed in the source domain affects the dynamic response in a manner that closely aligns with the damage effect in the target structure. For instance, if the damage on the target structure is material-specific, e.g., lamination in metal panels, considering materials in the similarity assessment will be required to identify an ideal source structure, from which it is possible to learn how this kind of damage affects the extracted features. On the other hand, the attributes related to the geometry and dimensions of the structure could be the main elements to be assessed in the case of very localised damage or damage-localisation tasks. Thus, the similarity assessment phase in PBSHM must be carefully analysed for the population to be assessed and may be case-specific to some extent. Therefore, it is necessary to analyse a comprehensive method, capable of taking into account multiple attributes for each node or edge of the AGs. This method should be capable of handling different types of attributes including discrete and continuous ones, and must be adaptable to the heterogeneous structures present in the database. The present section expands on GMN-based similarity assessment, introduced in [16], studying its ability to account for multiple attribute variations. Subsequently, a novel Euclidean or correlation-based similarity assessment metric is presented, and the results are discussed in Section 6.

GMNs have been proposed for similarity reasoning in [62,63], as a GNN extension [26] to directly deal with graph-structured data, because other methods would usually require the pre-processing of the graphs to transform them in simpler structures [64], leading to information losses. Conversely, GNNs possess the capability to learn graph representations from graph-structured data, thereby enhancing various supervised learning tasks that rely on such data [17]. These methods are based on Recursive Neural Networks (RNNs) [65], which are already able to process graphs, although with some limitations on the type of graph and task. Another strategy for dealing with graphs, which underlies GNN, is related to Markov-chain models [66]. In particular, these models include the random-walk theory, which is valuable for modelling events occurring randomly from a current state and where the possible states are represented by graphs [67]. From the extension of the random-walk concepts, the random-walk graph kernel was subsequently realised as a method for calculating similarity based on the behaviour of a simultaneous random walk on them [68]. An application of this method for PBSHM can be found in [20]. Instead, GNNs have been further developed mainly for improving supervised prediction problems by learning node representations and embeddings [69,70]. This machine-learning class captures graph patterns, considering each node, its neighbours and the associated information. GMNs join this family of methods, focussing on the similarity-learning problem [26], matching nodes across different graphs and identifying their differences to compute a similarity score with increased accuracy as a result of node and edge attribute embedding.

The workflow for computing a similarity index between a pair of graphs requires an encoding phase, a matching phase and an aggregation of the response into a quantitative similarity measure [26]. In the encoding phase, each AG is processed via multi-layer



(a) *No-winglets configuration.* (b) *End-winglets configuration.*

(c) *Middle-winglets/engines configuration.* (d) *Both-winglets/engines configuration.*

Fig. 11. The GARTEUR IE models in the four topological configurations: (a) no winglets are present on the wings, (b) winglets are present at the end of the wings, (c) winglets are in the middle of the wings to simulate the presence of an engine, and (d) winglets are in both the middle and end positions.

perceptrons (MLPs) to generate a vector representation of each node and edge that captures the main characteristics of the AG. Afterwards, the node representations are updated iteratively in a propagation layer. The novelty of GMNs is that the propagation is not only based on the local neighbourhood information but also on a cross-graph vector measuring the matching between the nodes of two graphs. After cross-graph matching, an aggregation module is adopted to produce the overall representation of each graph from the node representations, and a similarity metric is defined on the resulting vector space. The GMN is trained via pairs of labelled graphs (G1, G2, ℓ), where ℓ is a similarity measure, to learn how to embed the graphs into a vector space in which similar graphs are close to each other. Alternatively, the network can be trained using triplets of graphs (G1, G2, G3) and relative similarity measures. The training objective minimises the difference between the predicted similarity and the ground truth ℓ, or it ensures that G1 is closer to G2 than G3 in the embedding space. Further details on this method can be found in [26,63].

The use of GMNs in PBSHM requires building IE models and AGs for training the network, and several training examples would be necessary. Thus, synthetic models can be generated, as proposed in [16]. This dataset should represent the variability of the examined population. For instance, if multiple topological classes can be identified, they should be represented on the dataset using a balanced number of examples. Analogously, attribute variability should be represented by carefully building the synthetic models. These models can be used to train and validate the network. Subsequently, the GMN can be tested on a synthetic test dataset or the real members of the population.

In this study, the use of GMNs is extended to the population of laboratory-scale aircraft, which are represented by IE models. The only topological variation regards the presence of winglets, or engines, in four possible configurations (examples of the corresponding IE models are shown in Fig. 11).

The first topology class includes models without winglets, as in Fig. 2a, b, c, and d. The second class includes models with end winglets, as in Fig. 2f, g and h, and the third class includes models with additional elements to simulate the engine weight, as in Fig. 2(e). The last additional class considers both end winglets and middle winglets/engines, which could be obtained from the experimental models available. The construction of the GARTEUR IE model is described in [27], where its topology is compared with a broad population, including bridges and aircraft. However, to train the network for the similarity assessment task, it is necessary to provide a large dataset of labelled pairs of graphs. Thus, the first step of the GMN-based similarity assessment consists of building a synthetic dataset of 4800 IE models representing the GARTEUR population. These models are preliminarily distinguished into multiple similarity classes, depending on the four topological configurations, four scaling groups and the materials used. In addition, the attributes of materials, geometry, and element dimensions are randomly defined for each class, as shown in [28]. The synthetic dataset is divided into training, validation, and test sets. The validation set includes 1% of the IE models from the original dataset.

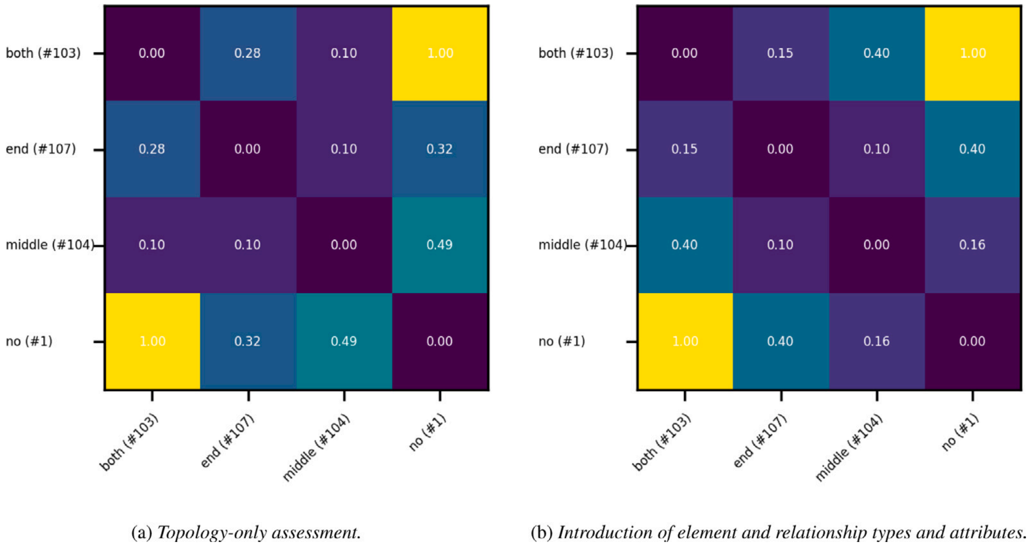


Fig. 12. Distance matrix of the aircraft population based on the four topology classes.

Instead, the test set includes 3.125% of the IE models, to observe the variation of the results as the attributes change. The GMN training is performed via pairs of GARTEUR IE models from the dataset, as shown in the Flow diagram of Fig. 1, and multiple similarity rules are implemented. Initially, the GMN is applied using just the four different topology classes for classifying similar and dissimilar models. It is implemented by defining a similarity rule which labels as similar all the pairs of graphs sharing the same topology class. Afterwards, the method is expanded to include the scaling groups as features to be examined for determining the similarity in the training dataset. Subsequently, the assessment is further expanded by considering the topology, dimensions (or scaling groups) and wing materials in the similarity rule. Finally, the trained network is tested on the experimental population. The results are shown in Section 5.1 in terms of normalised distances $\mathrm{d}_{i,j}$ between each pair of graphs (i, j). Subsequently, the GMN-based similarity can be expressed as $\zeta_{i,j} = 1 - \mathrm{d}_{i,j} \in [0, 1]$.

5.1. GMN results

The current section shows the results of the GMN-based similarity assessment. The first rule for assessing similarity only considers topology-based classes, such as aircraft with no winglets, with end winglets, with middle winglets or with engines, and with both winglets and engines. Aircraft belonging to the same class are considered similar. If they do not belong to the same class, they are considered dissimilar. The results of the comparison, when no attributes are embedded in the network, are presented in Fig. 12(a). The graph distance is normalised between 0 and 1. It is worth noting that the GMN is capable of distinguishing between different topology classes, with maximum distance values observed between graphs with no winglets and those with both winglets and engines. However, to enhance performance, it is necessary to introduce some attributes. Indeed, the study focuses on how the dimensions and the materials of test structures impact on the vibration-based features that can be experimentally extracted and used in knowledge transfer, exploring how these attributes can be embedded in graphs and integrated into the similarity rule used in GMN training. Therefore, categorical attributes such as material types, geometry types, and contextual labels are introduced for each graph node, following the PBSHM schema. Fig. 12(b) demonstrates how embedding element and relationship types, as well as the attributes, can greatly improve the ability to distinguish the middle-winglets/engines class from other topological classes. These attributes create a slight variation in distance values within each cluster. However, to evaluate the influence of dimensions, it is necessary to incorporate them into the classification rule used to label the graph pairs, as shown in [28]. The dataset is divided into four clusters based on the scaling factor of the IE model relative to the original model. By classifying graphs as similar when they display the same topology and scaling group, the distances presented in Fig. 13 are obtained. These are computed after 10 000 training steps.

The clustering of topological classes is less evident because of the simultaneous impact of scaling ratios. Consequently, the distance values between various topological classes are closer. Nevertheless, the matrix demonstrates an increase in distance values when the topology shifts from the both-winglets configuration to the no-winglets configuration. Furthermore, for each combination of topology classes, an increase in distance is noticeable as the scaling factor between the IE models grows. Subsequently, the GMN-based similarity assessment is extended to include the influence of the wing materials in the similarity rule. The same synthetic dataset is adopted during training. However, the training labels for each IE model pair (G1, G2) are set as unitary only when the graphs display the same topology, scaling group, and wing materials. Thus, the network can distinguish structures as they vary in dimensions and topological classes, as shown in extracts from the synthetic test set in Fig. 14. Different materials can

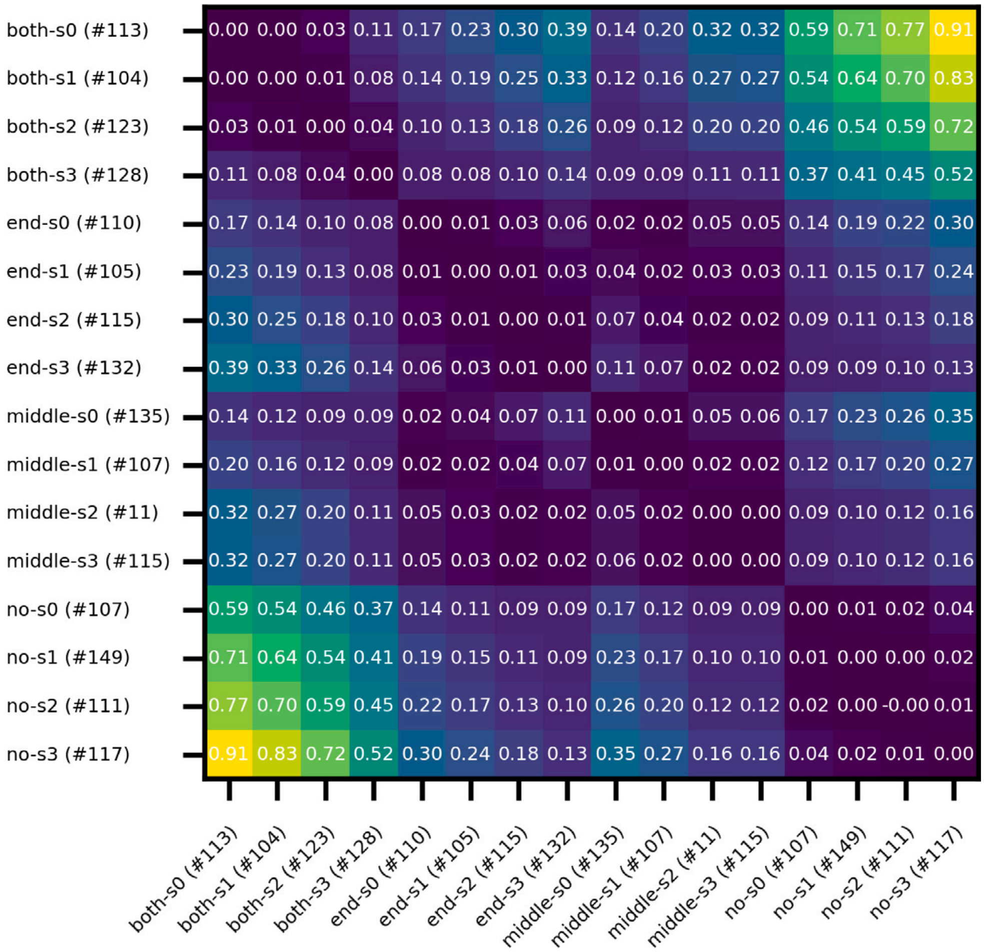
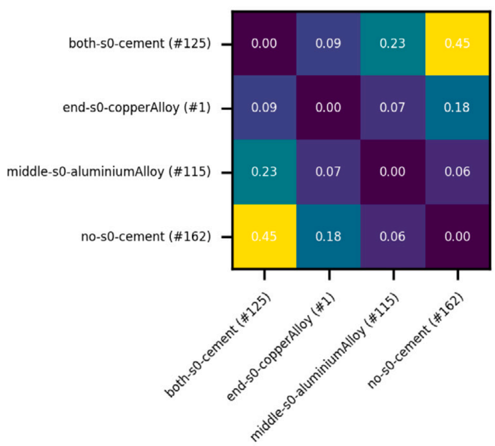
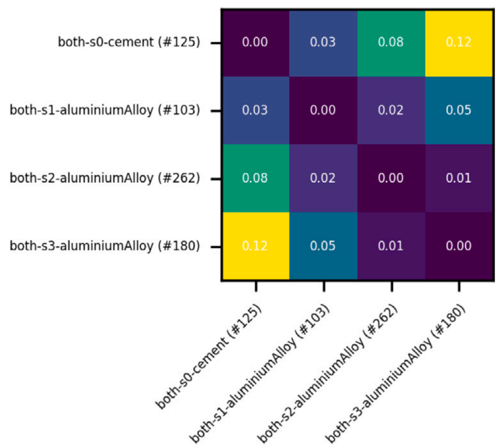


Fig. 13. Distance matrix of the aircraft population based on the 16 (topology and scaling group) classes, embedding element and relationship types, element contextual, material and geometry attributes. This is a subset of the results to focus on topology and scaling group variation.



(a) *Subset of the test set to highlight topological classes.*



(b) *Subset of the test set to highlight dimensions classes.*

Fig. 14. Distance matrix of the aircraft population based on the third similarity rule (topology + dimensions + materials), embedding element and relationship types, element contextual, material and geometry attributes.

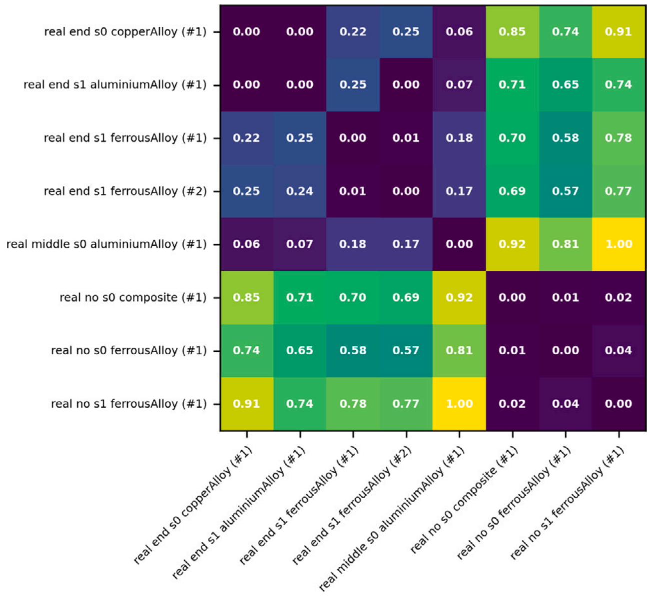


Fig. 15. Distance matrix of the real GARTEUR population based on the third similarity rule (topology + dimensions + materials), embedding element and relationship types, element contextual, material and geometry attributes.

be distinguished, albeit only categorically. Therefore, it is not possible to observe a continuous variation of similarity assessment results based on the actual mechanical properties of the materials used. Furthermore, the latter similarity rule is tested on the IE models created for the real members of the GARTEUR population. The similarity assessment results, computed after 20 000 training steps and normalised with $\ell \in [0, 1]$, are shown in Fig. 15. This Figure shows the confusion matrix of the distances measured within each combination of structures in the experimental population. Hence, the similarity between the graphs is defined as the complement of these distance measures. Given a pair of GARTEUR structures, it is possible to assess the distance between the respective AGs produced by the GMN, trained using the third similarity rule. Accordingly, a null distance is computed between a graph and itself. In contrast, an increase in the distance between two distinct graphs is exhibited in the presence of differences in their topology, dimensions, or materials. Two main clusters become more apparent, which relate specifically to topological classes, while the influence of attributes is present but bears minor relevance to overall results. Therefore, although some AGs in which there is only one element of variation are associated with each other, these results can be used as a basis for identifying cases in which better knowledge-sharing performance is expected. However, the results of this similarity assessment cannot be correlated with the overall outcomes of knowledge transfer, as shown in Fig. 16, which illustrates the damage-localisation performance as a function of the GMN-based similarity values. The analysis reveals two primary clusters driven by topological similarities, while other attributes exert only a secondary influence, failing to account for the complete variability domain-adaptation results. The insights derived from these results prompted a deeper investigation into the assessment of similarity, detailed in the following section.

6. Correlation-based similarity assessment

The definition of an adequate distance metric for PBSHM similarity assessment strictly depends on the characteristics of the population under analysis. Although the GMN-based assessment proved the ability to embed attributes and consider their variability for evaluating the distance between graphs, this type of approach prioritises the topology of the graphs. Therefore, the current study additionally explores a second metric based on the Euclidean distances between the aircraft attributes and topological classes [71]. This metric can be developed in such an application, as topological differences do not prevail over attribute differences, as seen from knowledge-transfer results. The properties of the Garteur population are summarised in Table 2. Four main elements distinguish the population of structures. These are the material density, the stiffness, the scale, and the topology. Therefore, the pairwise distance between two structures can be computed as a four-dimensional ℓ^2-norm Euclidean distance, as shown in Eq. (10), where s_i is a four-dimensional vector comprising numerical values to represent the aforementioned properties of the ith structure. This metric is chosen for its simplicity and interpretability, to provide a straightforward measure and capture point the distances between the attributes of source and target structures.

$$\mathrm{d}_{i,j} = \|s_i - s_j\| = \sqrt{\sum_k (s_i^k - s_j^k)^2} \tag{10}$$

In addition, by normalising the distance to the interval $\mathrm{d}_{i,j} \in [0, 1]$, the pairwise similarity metric can be expressed as $\zeta_{i,j} = 1 - \mathrm{d}_{i,j} \in [0, 1]$. The Euclidean distance calculation necessitates that the properties being compared are represented by numerical

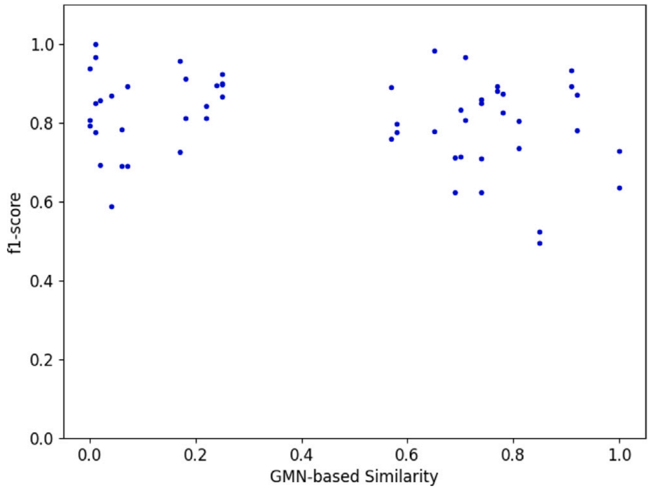


Fig. 16. Scatter plot of the F1 scores vs. the similarity results estimated using the GMN-based approach.

Table 2
Summary of the GARTEUR population's properties. Young's modulus is denoted by E and mass density is denoted by ρ.

	Material	ρ (kg/m^3)	E (GPa)	Scale	Topology
G1	Copper alloy	8400	90	s0	End winglets
G2	Aluminium	2710	68	s1	End winglets
G3	Steel	8000	200	s1	End winglets
G4	Steel	8000	200	s1	End winglets
G5	Aluminium	2710	68	s0	Middle engine
G6	Steel + composite	3000	250	s0	No winglets
G7	Steel	8000	200	s0	No winglets
G8	Steel	8000	200	s1	No winglets

values. Therefore, each of the attributes is transformed accordingly. The mechanical properties are normalised in the $[0, 2]$ range. An analogous approach is used for the scaling factor, associating 0 with the small-scale models and 2 with the large-scale models. Instead, the topologies are scaled according to their relative Jaccard similarity, computing the Jaccard index between simplified AGs, using only one element for each aircraft component. This preliminary re-scaling of the attributes and topological index guarantees that the properties are equally weighted in the initial Euclidean distance, ensures the comparability between the four properties and avoids biases introduced by attributes of different scales. Furthermore, each property can be associated with a weight w_k according to its sensitivity of the behaviour of the structure to the property.

$$\mathrm{d}_{i,j} = \|s_i - s_j\|_w = \sqrt{\sum_k w_k (s_i^k - s_j^k)^2} \tag{11}$$

The weights $\{w_1, w_2, w_3, w_4\}$ in Eq. (11) are set using an informed approach. This approach consists of selecting w_k by maximising the Pearson correlation coefficient r computed between the domain-adaptation results and the similarity ζ, under the constraint of $\|w\| = 1$. The considered knowledge-transfer performance refers to the damage-localisation task, and it is computed in terms of the F1 score (Fig. 8).

6.1. Correlation-based results

The resulting weights are $\{2.55 \times 10^{-01}, 2.51 \times 10^{-17}, 1.198 \times 10^{-01}, 9.59 \times 10^{-01}\}$, suggesting that Young's modulus does not significantly affect the transfer efficacy in the GARTEUR population. Thus, domain adaptation techniques guarantee optimal knowledge sharing despite variations in elastic modulus because they can harmonise the effect of these variations. Conversely, density, topology, and scaling have a higher impact on the effectiveness of knowledge transfer. The transfer performance trend against the results of the correlation-based similarity assessment is shown in Fig. 17, where the maximum performance coincides with the knowledge transfer between the two analogous models end-s1-ferrousAlloy (#1) and (#2), in which $\zeta_{i,j} = 1$. The application of this approach illustrates a higher level of consistency between the transfer results and the similarity than the GMN-based assessment shown in Fig. 16. Indeed, these graphs estimate the correlation between the two PBSHM phases: similarity assessment and knowledge transfer. Despite variability in the data compared to the median, especially for low similarity values, the knowledge-transfer performance monotonically increases as a function of the similarity and produces a positive Pearson coefficient of 0.365. Conversely, no correlation is observed in the results based on the GMN approach, which is attributed to the unbalanced influence

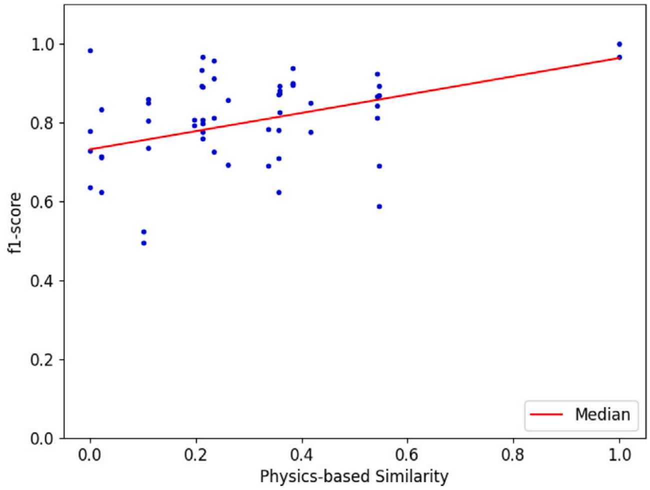


Fig. 17. Scatter plot of the F1 scores vs. the similarity assessment results estimated using the correlation-based approach.

of the topological clusters on the overall similarity assessment. Subsequently, these results may be used to evaluate the EVIT, as presented in [9,71] to optimise transfer learning.

7. Discussions and conclusions

The present study investigates the challenge of sharing knowledge within a heterogeneous population of laboratory-scale aircraft via transfer-learning approaches, specifically domain adaptation, to define the influence that different sources of heterogeneity, such as material properties, geometry, and dimensions, have on PBSHM performance. Therefore, one of the novelties of this analysis concerns the application of knowledge transfer to an entire population of structures, evaluating all possible combinations of source and target. In addition, PBSHM aims to determine which structures are appropriate for knowledge transfer to decrease negative transfer issues, particularly in heterogeneous populations. Thus, the current research explores efficient methods to consider attribute influence in the similarity assessment process and the interdependence of the phases of similarity assessment and knowledge transfer.

Each member of the population has been tested to extract its modal parameters under undamaged conditions and simulated damaged conditions, building a comprehensive dataset for validating the PBSHM approach. Subsequently, NCA is introduced as a preprocessing step to provide a preliminary feature alignment, and JDA is employed to harmonise each combination of source and target datasets to facilitate damage detection and localisation tasks. The domain-adaptation performance is examined across the population, proving the effectiveness of the PBSHM strategy. Specifically, all combinations of source and target structures lead to excellent damage-detection performance after domain adaptation. Instead, the damage-localisation outcomes are more affected by the attributes of the structures chosen for knowledge transfer. In this case, optimal results are achieved in clusters of similar structures, particularly in terms of topology. Lower performance is observed when there are differences in topology and attributes, such as geometry and material properties.

Moreover, two strategies for similarity assessment are proposed and compared to examine how to effectively evaluate the influence of material properties, geometric configurations, and dimensional variations. A GMN-based similarity assessment for heterogeneous structures has been performed. This assessment involves embedding the attributes into the GMN and integrating them into the similarity rules to improve the accuracy of similarity assessment and inform knowledge transfer into the PBSHM framework. The GMN has been trained on a synthetic training population and additionally tested on the AGs of the experimental population consisting of eight aircraft models, providing a significant overview regarding the issue of "when to transfer". Indeed, the GMN correctly associates the most similar models in two clusters and provides increasing distance with growing dimensional and material heterogeneity. However, two primary clusters, driven by topological similarities, have been identified. Additional attributes have a minor influence and fail to account for the variability observed in the domain adaptation results across the population.

Thus, the current study proposes a novel distance metric based on a weighted four-dimensional Euclidean distance. This distance metric considers the material density, stiffness, scaling, and topology of the GARTEUR structures. By implementing this similarity-assessment approach, a higher correlation with the results of knowledge transfer has been achieved, because the optimisation algorithm identifies the most relevant attributes in an Euclidean space. Hence, this methodology proves to be appropriate to study populations of heterogeneous structures where topological differences are less relevant than attribute variations, effectively facilitating knowledge transfer. Additionally, valuable insights are obtained regarding which sources are most relevant to ensure a positive transfer. Further developments of this study will regard the introduction of environmental variations in the features acquired from the aircraft population, to investigate its effect on the knowledge-sharing performance.

Table A.3
Natural frequencies extracted from the average FRFs of the population of experimental aircraft models.

Structure	Condition	f1 (Hz)	f2 (Hz)	f3 (Hz)	f4 (Hz)	f5 (Hz)	f6 (Hz)	f7 (Hz)
end-s0-copperAlloy (#1)	0-0	7.90	21.88	44.73	47.66	60.94	178.52	300.00
end-s0-copperAlloy (#1)	1-1	7.81	21.68	44.70	47.66	60.55	177.34	300.00
end-s0-copperAlloy (#1)	1-2	7.81	21.46	44.72	47.62	60.55	176.37	300.00
end-s0-copperAlloy (#1)	1-3	7.90	21.29	44.44	47.08	59.78	175.93	299.82
end-s0-copperAlloy (#1)	1-4	7.61	20.90	44.12	46.88	60.74	178.13	286.65
end-s0-copperAlloy (#1)	1-5	7.90	21.34	44.71	47.23	60.87	178.12	299.62
end-s0-copperAlloy (#1)	1-6	7.90	21.09	44.70	47.65	60.91	178.50	299.41
end-s1-aluminiumAlloy (#1)	0-0	6.24	15.71	36.58	36.86	39.97	48.10	63.63
end-s1-aluminiumAlloy (#1)	1-1	6.09	15.62	36.43	36.78	39.91	47.80	62.80
end-s1-aluminiumAlloy (#1)	1-2	6.15	15.68	36.49	36.80	39.94	47.80	63.51
end-s1-aluminiumAlloy (#1)	1-3	6.23	15.62	36.50	36.78	39.93	47.89	63.42
end-s1-aluminiumAlloy (#1)	1-4	6.24	15.38	36.00	36.71	39.95	48.06	63.62
end-s1-aluminiumAlloy (#1)	1-5	6.24	15.68	36.54	36.80	39.96	48.07	63.51
end-s1-aluminiumAlloy (#1)	1-6	6.23	15.68	36.58	36.86	39.96	48.08	63.57
end-s1-ferrousAlloy (#1)	0-0	5.64	14.59	29.83	32.39	32.80	38.70	60.53
end-s1-ferrousAlloy (#1)	1-1	5.62	14.44	28.54	32.38	32.69	37.45	60.10
end-s1-ferrousAlloy (#1)	1-2	5.63	14.37	28.87	32.38	32.78	38.14	60.32
end-s1-ferrousAlloy (#1)	1-3	5.64	14.48	29.71	32.38	32.79	38.52	60.02
end-s1-ferrousAlloy (#1)	1-4	5.63	13.43	28.95	32.38	32.77	38.66	58.83
end-s1-ferrousAlloy (#1)	1-5	5.64	14.46	29.74	32.39	32.79	38.63	60.08
end-s1-ferrousAlloy (#1)	1-6	5.64	14.46	29.71	32.39	32.79	38.57	60.24
end-s1-ferrousAlloy (#2)	0-0	6.09	16.09	35.91	37.53	40.81	46.93	62.40
end-s1-ferrousAlloy (#2)	1-1	5.99	15.90	35.65	37.11	40.52	46.71	62.05
end-s1-ferrousAlloy (#2)	1-2	6.02	16.00	35.75	37.33	40.71	46.81	62.06
end-s1-ferrousAlloy (#2)	1-3	6.08	16.02	35.86	37.27	40.75	46.89	62.26
end-s1-ferrousAlloy (#2)	1-4	6.05	15.40	35.70	37.26	40.72	46.92	61.88
end-s1-ferrousAlloy (#2)	1-5	6.07	16.00	35.87	37.48	40.81	46.90	62.30
end-s1-ferrousAlloy (#2)	1-6	6.08	15.99	35.88	37.48	40.80	46.89	62.32
middle-s0-aluminiumAlloy (#1)	0-0	33.60	45.15	94.67	126.41	181.11	287.19	317.50
middle-s0-aluminiumAlloy (#1)	1-1	32.04	44.44	93.19	126.21	179.35	284.66	315.33
middle-s0-aluminiumAlloy (#1)	1-2	32.85	44.79	94.17	126.22	180.92	284.65	313.70
middle-s0-aluminiumAlloy (#1)	1-3	33.60	44.94	93.59	125.22	178.77	286.71	316.56
middle-s0-aluminiumAlloy (#1)	1-4	33.60	42.03	93.90	119.82	181.11	286.57	316.38
middle-s0-aluminiumAlloy (#1)	1-5	33.59	44.93	94.56	125.02	180.92	286.82	317.50
middle-s0-aluminiumAlloy (#1)	1-6	33.51	44.94	94.56	125.83	173.68	286.62	308.88
no-s0-composite (#1)	0-0	36.93	55.88	175.64	203.86	217.84	247.15	280.56
no-s0-composite (#1)	1-1	34.26	53.14	172.18	202.90	215.25	243.24	273.13
no-s0-composite (#1)	1-2	35.41	54.11	174.47	203.37	216.86	243.63	277.23
no-s0-composite (#1)	1-3	36.73	55.68	172.89	202.80	215.63	245.16	278.97
no-s0-composite (#1)	1-4	36.93	53.12	173.28	203.15	217.45	246.02	279.74
no-s0-composite (#1)	1-5	36.93	55.69	175.25	203.85	217.44	246.36	279.94
no-s0-composite (#1)	1-6	36.74	55.69	175.45	203.85	217.70	247.01	280.07
no-s0-ferrousAlloy (#1)	0-0	33.60	64.28	129.73	169.15	185.84	223.12	291.69
no-s0-ferrousAlloy (#1)	1-1	31.46	62.52	124.84	162.56	179.55	213.15	283.29
no-s0-ferrousAlloy (#1)	1-2	32.61	63.30	128.56	165.35	184.24	213.16	286.42
no-s0-ferrousAlloy (#1)	1-3	33.59	63.70	124.45	162.94	180.33	223.13	291.50
no-s0-ferrousAlloy (#1)	1-4	33.57	59.19	127.78	166.47	185.22	223.35	291.69
no-s0-ferrousAlloy (#1)	1-5	33.60	63.69	129.19	168.30	185.41	222.28	290.65
no-s0-ferrousAlloy (#1)	1-6	33.02	63.69	128.94	169.01	183.26	220.57	287.59
no-s1-ferrousAlloy (#1)	0-0	8.09	16.84	34.66	34.69	34.69	45.61	64.55
no-s1-ferrousAlloy (#1)	1-1	7.44	16.24	33.10	33.10	33.10	43.52	63.14
no-s1-ferrousAlloy (#1)	1-2	7.75	16.44	34.54	34.54	34.54	45.39	63.83
no-s1-ferrousAlloy (#1)	1-3	7.98	16.68	33.07	33.07	33.07	43.75	64.17
no-s1-ferrousAlloy (#1)	1-4	7.97	16.00	33.73	33.73	33.73	45.00	62.97
no-s1-ferrousAlloy (#1)	1-5	7.66	16.41	34.51	34.51	34.51	44.91	64.19
no-s1-ferrousAlloy (#1)	1-6	7.88	16.66	34.60	34.60	34.60	44.88	64.05

CRediT authorship contribution statement

Giulia Delo: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Data curation, Conceptualization. **Aidan J. Hughes:** Writing – review & editing, Software, Methodology, Data curation, Conceptualization. **Cecilia Surace:** Writing – review & editing, Visualization, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Keith Worden:** Writing – review & editing, Visualization, Supervision, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

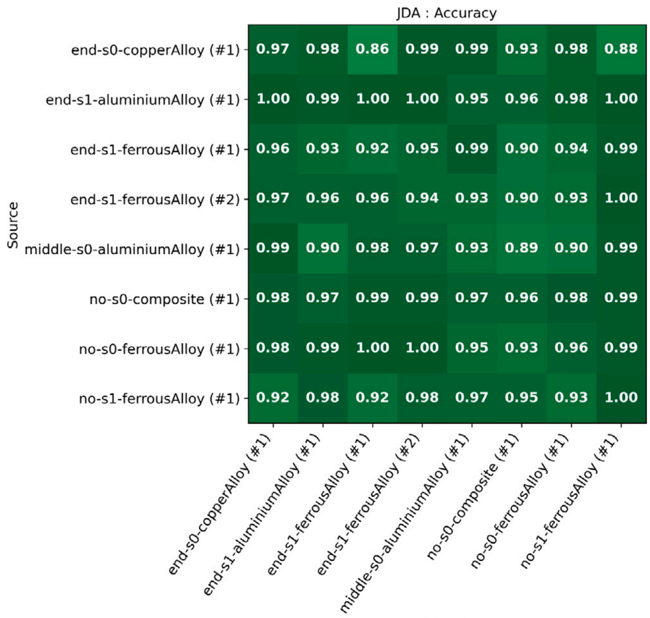


Fig. B.18. Damage-detection performance across the population: accuracy confusion matrix.

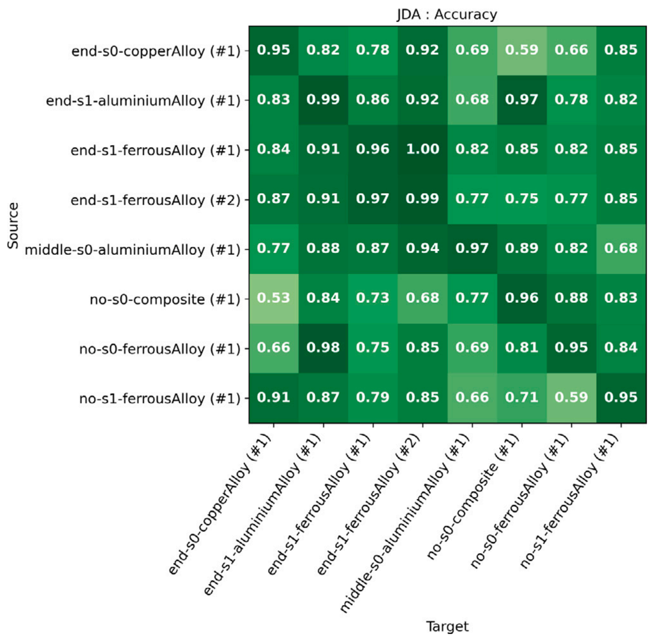


Fig. B.19. Damage-localisation performance across the population: accuracy confusion matrix.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Keith Worden reports financial support was provided by Engineering and Physical Sciences Research Council. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

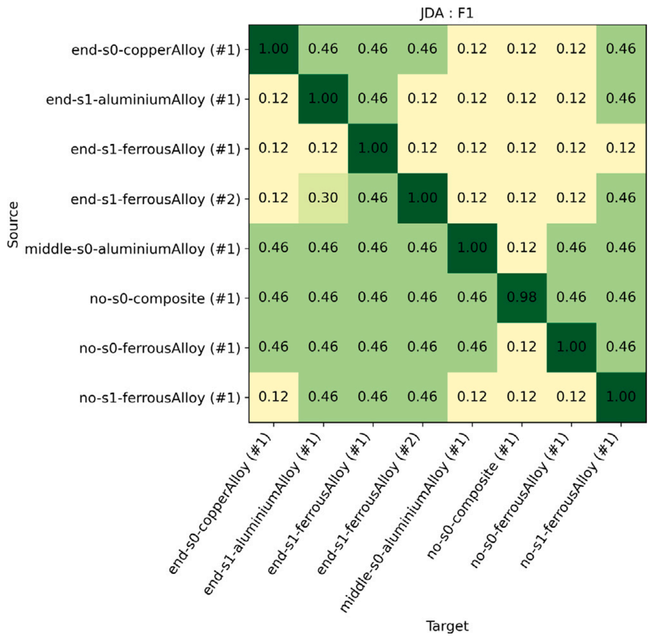


Fig. C.20. Damage-detection performance across the population without domain adaptation: F1 score confusion matrix.

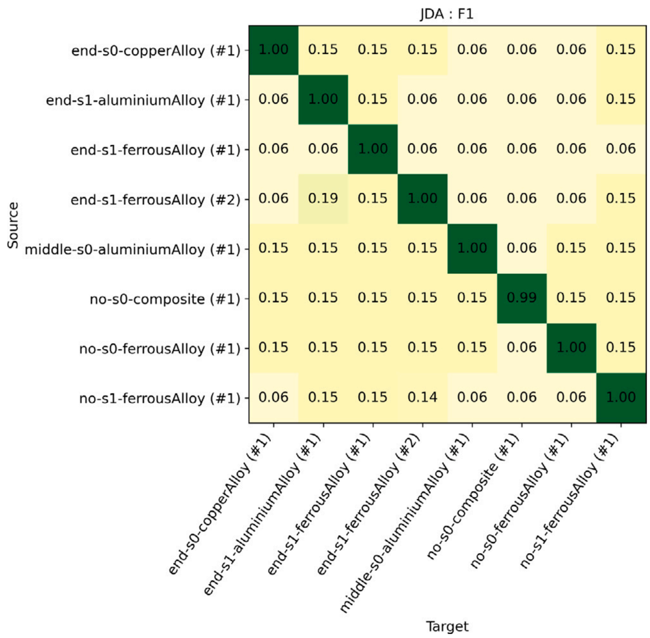


Fig. C.21. Damage-localisation performance across the population without domain adaptation: F1 score confusion matrix.

Acknowledgements

The authors of this paper gratefully acknowledge the support of the UK Engineering and Physical Sciences Research Council (EPSRC) via grant reference EP/W005816/1. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Appendix A. Experimental natural frequencies

See Table A.3.

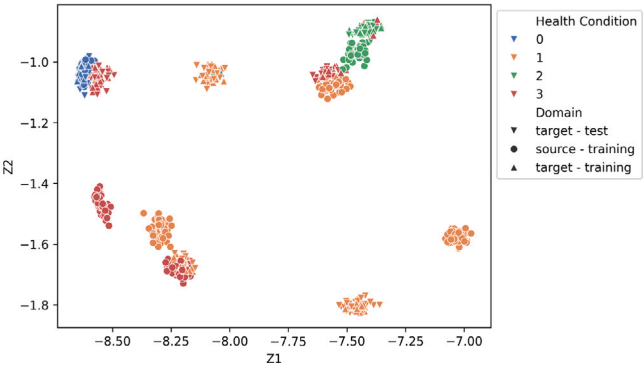


Fig. D.22. Transformed features in the latent space $Z = \in \mathbb{R}^{N x k}$, where N is the number of samples and k = 2. These features are computed for the damage-localisation task considering JDA on the no-s1-ferrousAlloy (#1) GARTEUR (source domain) and the middle-s0-aluminiumAlloy (#1) GARTEUR (target domain).

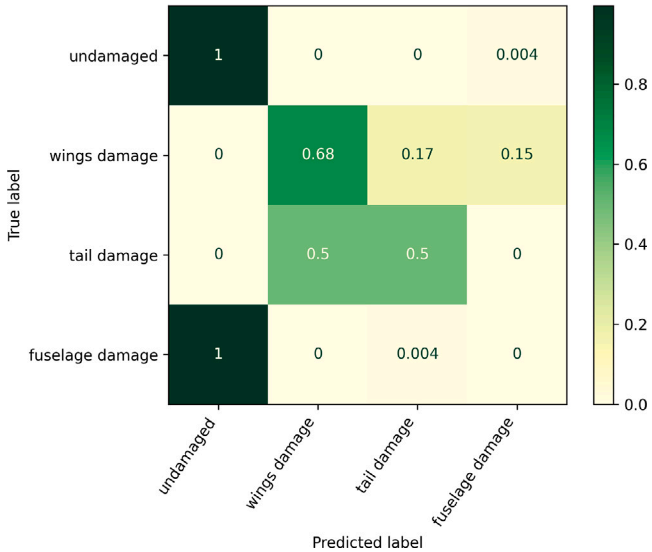


Fig. D.23. KNN confusion matrix for the damage-localisation task considering JDA on the no-s1-ferrousAlloy (#1) GARTEUR (source domain) and the middle-s0-aluminiumAlloy (#1) GARTEUR (target domain).

Appendix B. Overall JDA results (Accuracy)

See Figs. B.18 and B.19.

Appendix C. Overall results without domain adaptation (F1 score)

See Figs. C.20 and C.21.

Appendix D. Detailed results of the localisation task performed between the end-s1-ferrousalloy (#1) and the no-s0-composite (#1) structures

See Figs. D.22 and D.23.

Data availability

Data will be made available on request.

References

[1] C.R. Farrar, K. Worden, Structural Health Monitoring: A Machine Learning Perspective, John Wiley & Sons, 2012.
[2] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, D.J. Inman, Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks, J. Sound Vib. 388 (2017) 154–170.
[3] H. Khodabandehlou, G. Pekcan, M.S. Fadali, Vibration-based structural condition assessment using convolution neural networks, Struct. Control. Heal. Monit. 26 (2) (2019) e2308.
[4] C.S.N. Pathirage, J. Li, L. Li, H. Hao, W. Liu, P. Ni, Structural damage identification based on autoencoder neural networks and deep learning, Eng. Struct. 172 (2018) 13–28.
[5] L.A. Bull, P.A. Gardner, J. Gosliga, T.J. Rogers, N. Dervilis, E.J. Cross, E. Papatheou, A.E. Maguire, C. Campos, K. Worden, Foundations of population-based SHM, Part I: Homogeneous populations and forms, Mech. Syst. Signal Process. 148 (2021) 107141.
[6] J. Gosliga, P.A. Gardner, L.A. Bull, N. Dervilis, K. Worden, Foundations of population-based SHM, Part II: Heterogeneous populations–graphs, networks, and communities, Mech. Syst. Signal Process. 148 (2021) 107144.
[7] P.A. Gardner, L.A. Bull, J. Gosliga, N. Dervilis, K. Worden, Foundations of population-based SHM, Part III: Heterogeneous populations–mapping and transfer, Mech. Syst. Signal Process. 149 (2021) 107142.
[8] A. Rytter, Vibrational Based Inspection of Civil Engineering Structures (ph.D.-Thesis), 1993, defended publicly at the University of Aalborg, April 20, 1993 PDF for print: 206 pp.
[9] A.J. Hughes, J. Poole, N. Dervilis, P. Gardner, K. Worden, Quantifying the value of information transfer in population-based SHM, in: Proceedings of the 42nd International Conference on Modal Analysis, 2024, p. 14.
[10] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2009) 1345–1359.
[11] M.T. Rosenstein, Z. Marx, L.P. Kaelbling, T.G. Dietterich, To transfer or not to transfer, in: NIPS 2005 Workshop on Transfer Learning, Vol. 898, 2005, p. 4.
[12] S. Ben-David, R. Schuller, Exploiting task relatedness for multiple task learning, in: Learning Theory and Kernel Machines: 16th Annual Conference on Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003, Washington, DC, USA, August 24-27, 2003. Proceedings, Springer, 2003, pp. 567–580.
[13] B. Bakker, T. Heskes, Task clustering and gating for bayesian multitask learning, J. Mach. Learn. Res. 4 (2003) 83–99.
[14] S.-H. Cha, Comprehensive survey on distance/similarity measures between probability density functions, City 1 (2) (2007) 1.
[15] C.T. Wickramarachchi, E. Maguire, E.J. Cross, K. Worden, Measuring data similarity in population-based structural health monitoring using distance metrics, Struct. Heal. Monit. 23 (4) (2024) 2609–2635.
[16] D.S. Brennan, T.J. Rogers, E.J. Cross, K. Worden, Calculating structure similarity via a graph neural network in population-based structural health monitoring: Part II, in: Proceedings of the 40th IMAC, a Conference and Exposition on Structural Dynamics 2022, 2022, pp. 151–158.
[17] S. Ontañón, An overview of distance and similarity functions for structured data, Artif. Intell. Rev. 53 (7) (2020) 5309–5351.
[18] G. Levi, A note on the derivation of maximal common subgraphs of two directed or undirected graphs, Calcolo 9 (4) (1973) 341–352.
[19] S.V.N. Vishwanathan, N.N. Schraudolph, R. Kondor, K.M. Borgwardt, Graph kernels, J. Mach. Learn. Res. 11 (2010) 1201–1242.
[20] C.T. Wickramarachchi, J. Gosliga, E.J. Cross, K. Worden, On the use of graph kernels for assessing similarity of structures in population-based structural health monitoring, in: European Workshop on Structural Health Monitoring: EWSHM 2022-Volume 2, Springer, 2022, pp. 995–1004.
[21] K.M. Borgwardt, C.S. Ong, S. Schönauer, S. Vishwanathan, A.J. Smola, H.-P. Kriegel, Protein function prediction via graph kernels, Bioinform. 21 (suppl_1) (2005) i47–i56.
[22] S. Ontanón, E. Plaza, Similarity measures over refinement graphs, Mach. Learn. 87 (2012) 57–92.
[23] M. Neuhaus, H. Bunke, Automatic learning of cost functions for graph edit distance, Inform. Sci. 177 (1) (2007) 239–247.
[24] C. Garcia-Hernandez, A. Fernandez, F. Serratosa, Ligand-based virtual screening using graph edit distance as molecular similarity measure, J. Chem. Inf. Model. 59 (4) (2019) 1410–1421.
[25] P. Bille, A survey on tree edit distance and related problems, Theoret. Comput. Sci. 337 (1–3) (2005) 217–239.
[26] Y. Li, C. Gu, T. Dullien, O. Vinyals, P. Kohli, Graph matching networks for learning the similarity of graph structured objects, in: International Conference on Machine Learning, PMLR, 2019, pp. 3835–3845.
[27] G. Delo, A. Bunce, E. Cross, J. Gosliga, D. Hester, C. Surace, K. Worden, D. Brennan, When is a bridge not an aeroplane? Part II: A population of real structures, in: European Workshop on Structural Health Monitoring: EWSHM 2022-Volume 2, Springer, 2022, pp. 965–974.
[28] G. Delo, C. Surace, K. Worden, D.S. Brennan, On the influence of structural attributes for assessing similarity in population-based structural health monitoring, in: Proceedings of the 14th International Workshop on Structural Health Monitoring 2023: Designing SHM for Sustainability, Maintainability, and Reliability, DEStech Publications, Inc., 2023, p. 10.
[29] K. Weiss, T.M. Khoshgoftaar, D. Wang, A survey of transfer learning, J. Big Data 3 (1) (2016) 1–40.
[30] Q. Yang, Y. Zhang, W. Dai, S.J. Pan, Transfer Learning, Cambridge University Press, 2020.
[31] S. Dorafshan, R.J. Thomas, M. Maguire, Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete, Constr. Build. Mater. 186 (2018) 1031–1045.
[32] Y. Gao, K.M. Mosalam, Deep transfer learning for image-based structural damage recognition, Comput.-Aided Civ. Infrastruct. Eng. 33 (9) (2018) 748–768.
[33] D. Chakraborty, N. Kovvali, B. Chakraborty, A. Papandreou-Suppappola, A. Chattopadhyay, Structural damage detection with insufficient data using transfer learning techniques, in: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2011, Vol. 7981, SPIE, 2011, pp. 1175–1183.
[34] S.J. Pan, I.W. Tsang, J.T. Kwok, Q. Yang, Domain adaptation via transfer component analysis, IEEE Trans. Neural Netw. 22 (2) (2010) 199–210.
[35] P. Gardner, X. Liu, K. Worden, On the application of domain adaptation in structural health monitoring, Mech. Syst. Signal Process. 138 (2020) 106550.
[36] L. Souza, M.O. Yano, S. da Silva, E. Figueiredo, A comprehensive study on unsupervised transfer learning for structural health monitoring of bridges using joint distribution adaptation, Infrastructures 9 (8) (2024) 131.
[37] E. Figueiredo, J. Brownjohn, Three decades of statistical pattern recognition paradigm for SHM of bridges, Struct. Heal. Monit. 21 (6) (2022) 3018–3054.
[38] V. Giglioni, J. Poole, I. Venanzi, F. Ubertini, K. Worden, A domain adaptation approach to damage classification with an application to bridge monitoring, Mech. Syst. Signal Process. 209 (2024) 111135.
[39] G. Michau, O. Fink, Unsupervised transfer learning for anomaly detection: Application to complementary operating condition transfer, Knowl.-Based Syst. 216 (2021) 106816.
[40] S. van Baars Buisman, G. Michau, A. Rene, O. Fink, Domain adaptations for guided wave SHM of composites: Towards fleet monitoring, in: Proceedings of the 6th European Conference of the Prognostics and Health Management Society 2021, Vol. 6, Prognostics and Health Management Society, 2021, pp. 430–438.
[41] X. Zhou, C. Sbarufatti, M. Giglio, L. Dong, A fuzzy-set-based joint distribution adaptation method for regression and its application to online damage quantification for structural digital twin, Mech. Syst. Signal Process. 191 (2023) 110164.
[42] P. Gardner, L. Bull, J. Gosliga, J. Poole, N. Dervilis, K. Worden, A population-based SHM methodology for heterogeneous structures: Transferring damage localisation knowledge between different aircraft wings, Mech. Syst. Signal Process. 172 (2022) 108918.

[43] L. Bull, P. Gardner, N. Dervilis, E. Papatheou, M. Haywood-Alexander, R. Mills, K. Worden, On the transfer of damage detectors between structures: An experimental case study, J. Sound Vib. 501 (2021) 116072.
[44] A. Crocetti, L. Scussolini, G. Miraglia, R. Ceravolo, et al., Multi-class damage detection on experimental frames through transfer component analysis, E-J. Nondestruct. Test. 29 (2024) 1–8.
[45] G. Delo, R. Roy, K. Worden, C. Surace, On the use of the inverse finite element method to enhance knowledge sharing in population-based structural health monitoring, Comput. Struct. 307 (2025) 107635.
[46] J. Poole, P. Gardner, N. Dervilis, L. Bull, K. Worden, On statistic alignment for domain adaptation in structural health monitoring, Struct. Heal. Monit. 22 (3) (2023) 1581–1600.
[47] G. Delo, C. Surace, K. Worden, Knowledge sharing for improving damage identification across a population of heterogeneous laboratory-scale aircraft models, in: 31st International Conference on Noise and Vibration Engineering, ISMA 2024, Katholieke Universiteit Leuven Faculty of Engineering Department of Mechanical Engineering Division of Mechatronische Systeem Dynamica (LMSD), 2024, p. 11.
[48] E. Balmes, J.R. Wright, GARTEUR group on ground vibration testing: Results from the test of a single structure by 12 laboratories in europe, in: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 80401, American Society of Mechanical Engineers, 1997, V01AT03A004.
[49] M. Link, M. Friswell, Working group 1: Generation of validated structural dynamic models—results of a benchmark study utilising the GARTEUR SM-AG19 test-bed, Mech. Syst. Signal Process. 17 (1) (2003) 9–20.
[50] G. Delo, M. Mattone, C. Surace, K. Worden, Novelty detection across a small population of real structures: A negative-selection approach, in: XII International Conference on Structural Dynamics, 2023, p. 10.
[51] G. Delo, R. Roy, K. Worden, C. Surace, Using the inverse finite-element method to harmonise classical modal analysis with fibre-optic strain data for robust population-based structural health monitoring, Strain e12481 (2024) http://dx.doi.org/10.1111/str.12481.
[52] I. Behmanesh, B. Moaveni, Probabilistic identification of simulated damage on the Dowling Hall footbridge through Bayesian finite element model updating, Struct. Control. Heal. Monit. 22 (3) (2015) 463–483.
[53] E. Papatheou, G. Manson, R.J. Barthorpe, K. Worden, The use of pseudo-faults for novelty detection in SHM, J. Sound Vib. 329 (12) (2010) 2349–2366.
[54] K. Zaletelj, T. Bregar, D. Gorjup, J. Slavič, Ladisk/pyEMA: v0.24, 2020, http://dx.doi.org/10.5281/zenodo.4016671.
[55] G. Delo, D.S. Brennan, C. Surace, K. Worden, On the influence of structural attributes for transferring knowledge in population-based structural health monitoring, in: Proceedings of the 42nd International Conference on Modal Analysis, IMAC-XLII, 2024, p. 8.
[56] A. Arnold, R. Nallapati, W.W. Cohen, A comparative study of methods for transductive transfer learning, in: Seventh IEEE International Conference on Data Mining Workshops, ICDMW 2007, IEEE, 2007, pp. 77–82.
[57] M. Long, J. Wang, G. Ding, J. Sun, P.S. Yu, Transfer feature learning with joint distribution adaptation, in: 2013 IEEE International Conference on Computer Vision, 2013, pp. 2200–2207, http://dx.doi.org/10.1109/ICCV.2013.274.
[58] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf, A. Smola, A kernel method for the two-sample-problem, Adv. Neural Inf. Process. Syst. 19 (2006).
[59] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf, A. Smola, A kernel two-sample test, J. Mach. Learn. Res. 13 (1) (2012) 723–773.
[60] D.J. Hand, H. Mannila, P. Smyth, Principles of Data Mining, The MIT Press, 2001.
[61] C. Bishop, Pattern recognition and machine learning, Springer Google Sch. 2 (2006) 35–42.
[62] M. Gori, G. Monfardini, F. Scarselli, A new model for learning in graph domains, in: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005, Vol. 2, IEEE, 2005, pp. 729–734.
[63] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, IEEE Trans. Neural Netw. 20 (1) (2008) 61–80.
[64] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall PTR, 1998.
[65] A. Sperduti, A. Starita, Supervised neural networks for the classification of structures, IEEE Trans. Neural Netw. 8 (3) (1997) 714–735.
[66] W.-K. Ching, M.K. Ng, Markov chains, Model. Algorithms Appl. (2006).
[67] G.H. Weiss, R.J. Rubin, Random walks: theory and selected applications, Adv. Chem. Phys. 52 (1983) 363–505.
[68] U. Kang, H. Tong, J. Sun, Fast random walk graph kernel, in: Proceedings of the 2012 SIAM International Conference on Data Mining, SIAM, 2012, pp. 828–838.
[69] G. Katsimpras, G. Paliouras, Improving Graph Neural Networks by combining active learning with self-training, Data Min. Knowl. Discov. (2023) 1–18.
[70] P. Battaglia, R. Pascanu, M. Lai, D. Jimenez Rezende, et al., Interaction networks for learning about objects, relations and physics, Adv. Neural Inf. Process. Syst. 29 (2016).
[71] A.J. Hughes, G. Delo, J. Poole, N. Dervilis, K. Worden, Quantifying the value of positive transfer: An experimental case study, in: Proceedings of the 11th European Workshop on Structural Health Monitoring, EWSHM2024, 2024, p. 13.